ERIODICAL ROOM
ENERAL LIBRARY
NIV. OF MICH.

OCT 28 1930

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

HAROLD RUGG, *Editor*

VOL. XXI, No. 7 OCTOBER, 1930 $5.00 per year

CONTENTS

SHALL "HABITS" BE DISCARDED IN EDUCATIONAL PSYCHOLOGY? *William F. Bruce* 479

GENERALIZATION OF BRIGHT AND DULL CHILDREN, A COMPARATIVE STUDY WITH SPECIAL REFERENCE TO SPELLING. *Herbert A. Carroll*. 489

PHONIC VERSUS LOOK-AND-SAY TRAINING IN BEGINNING READING. *Raymond M. Mosher and Sidney M. Newhall*. 500

A COMPARISON OF DIFFICULTY AND ACCURACY IN THE LEARNING OF BRIGHT AND DULL CHILDREN IN A MOTOR-MEMORY TASK. *Frank T. Wilson* 507

FACTORS DETERMINING RELIABILITY OF TEST NORMS. *E. F. Lindquist* 512

HOW SUCCESSFUL COLLEGE STUDENTS STUDY. *James D. Weinland*. 521

THE POETRY OF PROMISE. *Rose Susan Malmud* 527

ON BITING VERSUS BARKING CRITICISM. *Walter F. Dearborn* . 545

ON THE RELIABILITY OF A STANDARD SCORE. *John W. Dickey* . 547

THE PROBABLE ERROR OF THE ACCOMPLISHMENT QUOTIENT. *C. L. Huffaker*. 550

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 552

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

Editor: HAROLD RUGG

Board of Associate Editors

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — E. L. THORNDIKE
FRANK N. FREEMAN — V. A. C. HENMON — L. M. TERMAN — GOODWIN B. WATSON

Assistant to the Editor: FRANCES M. FOSTER

The Journal is published monthly, except June to August, by Warwick and York, Inc., Baltimore, Md., and York, Pa.

The price is $5.00 a year, payable in advance; foreign postage 30 cents extra. Single current issues are 70 cents each, and subscriptions for less than a year are charged at single issue rate for each number ordered. Back volumes are $6.00 each; single back numbers, 90 cents each, except, when more than three years old, the price is $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence concerning editorial matters should be addressed to Harold Rugg, Editor, 425 West 123rd Street, New York City. Subscriptions should be sent to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

TEACHERS' HEADQUARTERS

European

Plan

Entirely

Fire-Proof

Centrally

Located

EDWARD DAVIS

Manager

HOTEL RENNERT

BALTIMORE, MARYLAND

The Psychology of Learning

1928 Revision — BY WILLIAM H. PYLE

$2.20 + 12¢ postage

Warwick & York, Inc., *Baltimore, Maryland*

The Acid Test

What more exacting test can there be of dictionary leadership than the judgment of the educational world?

Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.

Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.

The New International has been universally accepted by those best fitted to judge.

It is in accord with the best modern practice.

Use it in your own schools.

Write for our free booklets of interesting ways to teach the use of the dictionary.

GET THE BEST

G. & C. MERRIAM COMPANY

SPRINGFIELD, MASS.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXI | October, 1930 | Number 7

SHALL "HABITS" BE DISCARDED IN EDUCATIONAL PSYCHOLOGY?

WILLIAM F. BRUCE

State Normal School, Oneonta, N. Y.

The trail of progress in any science can be traced by noting the popular beliefs which have been dropped along the way. Throwing aside "common-sense" notions is a favorite pastime of those who do research in uncharted fields. Many educational psychologists have practically eliminated the word "mind" from their working vocabularies, and only faint traces of the ancient "instinct" concept remain. One of the chief reasons given for omitting these words is that such terms mislead the student due to their common connotations. Since the large packages, called "mind" and "instinct," have been dropped in the interest of upward flight, we venture to raise the question whether or not the whole bundle of "habits" should be discarded also.

Before taking this drastic step, however, let us be aware of a danger involved, and let us remember an alternative which is open. The danger is an old one, familiarly expressed as "throwing out the baby with the bath." While bent on eliminating "mind," as a "thing" or "mental substance," some critics have dropped at the same time all recognition of intelligence in behavior. Likewise, it has proved difficult to discard the separate, definite instincts without losing sight of the profound influence which may be credited legitimately to the pre-birth history of each individual. Being forewarned of this danger we may carefully select and retain any valid ideas carried in "habit."

Besides, it may be more profitable to take the alternative of redefining the old term. Those who still speak of "mind" and "instinct" have constructed new definitions for these time-honored words. Our terms are constantly undergoing redefinition, and if attention is turned specifically to this process a new meaning

may be conferred upon the symbol, "habit," which will eradicate the popular, pre-scientific misconceptions. At the same time those characteristics which withstand rigorous criticism may be conserved. Whether it is more enlightening to discard the term "habit," in the teaching of educational psychology, than to attempt a constructive alteration of the student's conception of habitual activity, is the question.

The discussion of "habit" in educational psychologies is frequently a source of confusion. Instead of clarifying the student's view of the learning process, the account given often produces a conflict of ideas, blocking his escape from the notions about habit which he has picked up in the home, street and church. Is it not confusing to find that habitual action carries such contrasting characteristics as "fixity" and "adjustability"? When one reads definitions of habit in recent texts, he usually finds that the author is following the thought of William James in referring to habit as "an automatic, ready-made response."[1] The "ready-made response" of James often becomes "*fixed* ways of reacting to particular situations."[2] But when the student turns to the psychologist's *description* of a typical habit, he may discover —if the definition has not obscured his vision—that *adjustability* is being emphasized.

Professor Judd's "Psychology of Secondary Education" presents a striking instance of evident inconsistency between a definition in which "inflexibility" is set down as an essential feature of habit, and a description in which readjustment predominates. Judd writes, "Habits are, in general, inflexible and unprogressive. They cannot be readjusted because of their fixity . . . "[3] On the following page the author describes the skillful use of hand tools as exemplifying habit. Note the range of flexible activity found in the habitual skill of the sawyer.

> He is instantly *responsive* to the sensations which come from the saw. In such a phase of his movement as following a line or making a straight cut, the expert is *guided* by sensations in the palm of the hand. Let the saw swerve ever so little, and the skillful workman makes the necessary *turn* of his hand. He knows, further, how to *adjust* his stroke to the different kinds of material; and he knows also that, when the board is very nearly divided, he must make a *skillful* stroke in completing the act.[4]

[1] James, W.: "Talks to Teachers." 1899, p. 66.
[2] Benson, Lough, Skinner, and West: "Psychology for Teachers." 1926, p. 65.
[3] Judd, C. H.: "Psychology of Secondary Education." 1927, p. 301.
[4] *Ibid.*, p. 302. (Italics not in the original.)

In this description adjustability, rather than fixity, is outstanding. In the final clause the word "skillful" is used to express that neat adjustment which is so characteristic of a well-formed habit. The value of this habit seems to depend upon its responsiveness to even the slightest change in the conditions. When the student of education reaches the conclusion that habitual action is characterized by responsiveness, he may be in a quandary concerning the commonly accepted definitions of habit. If adaptability is an essential feature of habits, why do definitions stress inflexibility?

There *is* a degree of stability in habits which must be recognized in description as well as in definition. It is highly probable that the expert sawyer of today will saw with a like skill tomorrow, and a year hence. We cannot avoid seeing the continuity in form of action so evident in hand skills and other forms of habitual activity. Habits do persist. But this persistent similarity, which appears when today's behavior is compared with that of yesterday, should not lead to the use of terms that exclude from view the wide range of responsiveness within many a habit. A student is quite likely to be misled by a writer who defines habit formation as learning "to act or to tend to act in the *same* way as one has done in the past."[1] It is far better to remember, as James pointed out long ago, that no one ever performs an act in exactly the same way twice.[2] In the light of the changes constantly occurring in the organism and its environment, "fixed" and "inflexible" appear to be misnomers for a kind of continuity, persistence and stability which involves adjustability as an essential characteristic. The redefinition of habit calls for different adjectives.

At this point it may prove helpful to borrow the phrase which Lloyd Morgan has used in redefining instinctive activity—"persistency with varied effort." According to his interpretation, when we say a pair of robins build a nest "by instinct," we do not imply a fixed series of actions previously laid down in the neuromuscular organization of the birds. Instead, the robins "instinctively" use the available materials in various ways suited to a particular nest-building situation. What really occurs, as a basis for the label "instinctive," is that the robins persist toward a *result* which is similar enough to the usual product that it is readily recognized as a robin's nest. But two nests are never exactly the same in construction. Typically, instinctive

[1] Benson, Lough, Skinner, and West: *Op. cit.*, p. 66.

[2] *Cf.* James, W.: "Psychology, Briefer Course." P. 154.

persistence does not employ exactly the same actions in any two cases. In general, such persistence involves a variety of efforts *adapted* to the peculiar conditions. Morgan's phrase provides an escape from the stability-adjustability conflict, which has been so confusing in the study of both "instinct" and "habit." When persistency is conceived as a movement toward a broadly perceived outcome, sufficient room is left for the adaptive effort needed within any instinct or any habit.

This conception of persistency shifts the emphasis from likeness in *specific* movements to likeness in the *general* character of the product. Consequently, the student's field of attention is widened so he will give adequate consideration to the "outer" consequences in the environment, without neglecting the "inner" conditions of the organism. In illustrating this view of habit, Thayer has used baseball pitching to show that "The sameness of an act applies to the outer result rather than to the activities involved in producing the outcome. The ball pitcher who is 'steady' is one who meets the ever-varying situations of the game in such a way that the same curves and speed are produced under differing circumstances."[1] The "steady pitcher" persists, in spite of differing conditions, toward results which appear to the catcher as "the same old curves," but which actually involve in their production a variety of rapidly changing readjustments of emotional impulses and sense perceptions which have never before been experienced. When steadiness, stability and persistence are understood in this way, these words carry the valid side of what James, Judd, and other psychologists may have meant when they used such terms as "automatic," "sameness," "fixity," in the definition of habit. But the student of education will find it safer to apply Morgan's phrase—"persistency with varied effort"—to habitual action, for this permits him to retain the sound idea of stability without being bound to the notion that responses can be "ready-made." Instead of accepting the statement, "Habits may become as fixed as instincts,"[2] the student will conclude, "Habits may involve as varied effort as do instincts, for they likewise persist toward similarity in the general result."

It will occur to the reader, no doubt, that some habits are so narrow in scope that little room is left for flexibility. A motorist may have "a habit of shifting gears," which is relatively "automatic"

[1] Thayer, V. T.: "The Passing of the Recitation." 1927, p. 105.

[2] Benson, Lough, Skinner, and West: *Op. cit.*, p. 66.

and "inflexible," and which, therefore, causes trouble when he tries to drive an auto with a different gear shift. In this case the similarity does seem to lie in the specific movements rather than in the general result. But the same motorist may have also "a habit of critical thinking," which is near the acme of flexibility. Such examples raise again the question how habits shall be defined, and they illustrate another source of confusion—that is, the wide contrast in the activities included under the term. The whole problem of the relation between narrow "fixed" habits and broad "flexible" ones emerges, and points on to the difficulty of reconciling "habit" and "intelligence." We cannot deal with these questions here, but they serve to emphasize the necessity of prying beneath the surface of any activity called a habit, in order to discover its specific characteristics and the conditions which produce it. When a term is used so broadly and loosely it is certainly not a dependable guidepost for the student who is seeking a more exact and scientific comprehension of the learning process. Therefore, instead of accepting habit as a symbol of a certain kind of activity, or of a particular causal factor, the student will do well to take it as a sign to look sharply into the *whole* matter. Let us turn, then, to a fundamental problem which arises in connection with that class of habits wherein flexibility in action accompanies stability in outcome.

When the conclusion is reached that adjustability of activity is an essential characteristic of efficient habits, the question naturally follows: What causes or conditions lead to the high degree of similarity often observed in the result? If our habits are constantly readjusting themselves why does so much continuity appear that it demands the use of this name? Why do flexible habits persist at all?

In attempting an answer to this question the student must be on his guard against "explanations" which narrowly limit the number of "causes" or "conditions" included. The prevalent explanations of the continuity observed in habitual action concentrate on the identification of a habit with the person who "has" it. The positive support which must be given by the environment, when the similarity lies in the result, is neglected. Furthermore, psychologies frequently so overemphasize the "establishment of neural bonds" as to give the student a notion that habit formation consists in a rather permanent alteration in a relatively restricted and separate section of the neuro-muscular equipment. Such an interpretation reminds the writer of an "explanation" once given him in grade school that the tides were

"caused" by the moon. When one finds, for instance, that at the Atlantic end of the Panama Canal there is only one high tide a day while at the Pacific end of the Canal there are two high tides daily, he begins to suspect either that the moon has peculiar powers or that there are other factors to be taken into account. And when the geography student learns, further, that the Pacific tides are frequently more than fifteen times as high as the Atlantic tides, at these two points little more than thirty miles apart, he may come to the conclusion that a number of "causes" must be investigated before the flow of the tides can be "explained." In like manner, the psychology student who observes habits carefully will give up the notion that they "belong" so exclusively to an individual, or that they are established within a limited area of the neural system. The persistence toward similarity in outcome, which non-critical common speech has labeled "habitual," cannot be explained scientifically without bringing to light a wide variety of conditioning factors.

For example, let us consider a habit which is well-formed and persistent in most of us—handwriting. Certainly an important factor which contributes to the continuance of the skill is the more or less flexible pattern of activity which has taken form during previous writing. Without doubt, this organization of activity has involved modification of the neural and muscular conditions in the hand, arm and brain. But it is also highly probable that activities which contribute to, or rather, are involved in the writing habit are continually occurring throughout the body as this habit is being "used." One condition, aiding in the maintenance of similarity, is that the writer keeps his *eyes* on the *result*. The *ears* also help in the detection of undesirable variations, such as a scratching pen. Nor will the behaviorists allow us to omit certain changes in *visceral* conditions which enter into that complexity of feeling and action called a habit. So it is generally agreed (though by no means adequately emphasized in many psychologies) that the formation and stability of a single habit depend upon the *whole* body of the man. Illness or old age, by altering the general health, often brings also a "shaky hand." But even this view does not carry the history of a habit outside of what we call the "individual." The habit may still be regarded as merely *in him*. On this narrow basis, the student is likely to assume that the writing habit becomes fixed in the school pupil, after a few years practice, as a phase of his physical behavior. But an adequate understanding requires that the "external" or environmental factors,

which are involved in handwriting, be considered along with the widespread "internal" reorganization which occurs.

The intimate relation of habitual action to environmental conditions can scarcely be appreciated except through noting the changes in action which accompany changes in these factors. The quality of handwriting depends, in part, upon certain equipment, such as well-made desks. If the pupil, on leaving school, takes a job where his writing is done on a pad held in the hand, numerous changes occur in his writing habit. It will not persist in quite the same form. Likewise, if a very rapid rate of writing is required the handwriting may be "ruined." When two such factors as speed and equipment are considerably altered the change in the handwriting may be quite as notable as its stability. The "natural" persistence appears to be greatly weakened by the lack of environmental support. But the reader may object to this interpretation on the ground that "abnormal" conditions have been introduced. He may assert: "Under *normal* conditions the handwriting habit does persist." This amounts to acknowledging, however, that the "external" conditions are extremely important factors in determining the degree of similarity which will appear in the results. Indeed, if "normal conditions" are to be accepted as the indispensable accompaniment of a habit which endures, is not this very much like saying that an organism acts in a similar way when confronted by similar conditions? If "habit" involves nothing more than the adjustment of an organism's activity in accordance with its environmental situation, it is time, certainly, to abandon the term as a meaningless one. If stability in results is to be linked invariably with stability in the environment—with "normal conditions"—the sooner the term is cast aside the better. For it is more appropriate to apply the name "adjustment" than "habit" to an action which fits so neatly into the situation.

But, in opposition to an interpretation of stability which would give so much credit to similarities in the environment, it may be maintained that the unique characteristic of a habit is the persistence of certain essential features of the activity *in spite* of changes in some "external" conditions. The handwriting may still be legible, even when comparatively defective equipment is used at the intense speed demanded by business service. The conditioning factor which dominates the situation is the *idea* of a certain result to be maintained, namely, legibility. The habit's stability depends in a fundamental way upon the aim or *ideal* of the writer. Applying this widely,

habitual action does not imply the mere *repetition* of activity by neurones and muscles which are "ready" to act because they have been "exercised frequently." Rather, habit involves a more or less continuous reorganization of the whole brain and bodily action for the *purpose* of maintaining a *result* which is foreseen and desired.

As Dewey says:

> Repetition is in no sense the essence of habit. Tendency to repeat acts is an incident of many habits but not of all . . . The essence of habit is an acquired predisposition to *ways* or modes of response, not to particular acts except as, under special conditions, these express a way of behaving. Habit means special sensitiveness or accessibility to certain classes of stimuli, standing predilections and aversions, rather than bare recurrence of specific acts.[1]

The stability of an effective, flexible habit centers in the "standing predilections" or purpose or ideal which the person holds. The results are continually checked against the ideal. With his ideal as a guide the individual continues to use his environmental equipment and "himself" to produce a similar outcome from time to time in spite of many alterations in conditions. Since the stability of such habits lies in the similarity of results rather than in the duplication of actions, and since results must be foreseen in order that they may be attained, it appears that aim, purpose or ideal constitutes the key to these habitual actions.

Ideals, however, are exposed to the possibilities of variation, quite as are other conditioning factors. For example, many pupils let their writing and their grammar "slide" when they escape from the pressure of the grade-school teacher. Other students continue to improve their "hand" and their command of English, because they look forward to social and professional benefits therefrom. These habits grow "stronger" in the sense that they become more effective as they approach the "ideal." But if an ideal is subject to change, for the better or the worse, or if it may be abandoned altogether, the assumed stability of flexible habits seems to rest upon an insecure foundation. Does this mean that a habit which is dominated by an ideal may be as unstable as a whim?

While it is true that ideals do change, and that in extreme cases this may occur "overnight," it is quite clear that the ideals which guide our habits in producing stable outcomes rest upon a foundation, broad and deep. Having made ideal the keynote of habit, we must not take the ideal as a narrow, limited "cause." Instead, we must go

[1] Dewey, John: "Human Nature and Conduct." 1922, p. 42.

back over the ground and at least suggest the multitude of factors which enter into making the ideal itself *live on*. In the first place, the social environment plays a large part in determining one's ideals. And the social situation is extremely stable in many of its aspects. To return to our illustration, the social demand for writing ability is persistent. Consequently, within a civilized community one finds not only many *reasons* for writing, but *equipment* is readily accessible also. The living physical body enters in by providing energy which must overflow in action. And this action flows, as far as may be, in accordance with the patterns which have been found fitting in previous situations. This means not mere similarity in physical movement, but also that a person is apt to "see" or "take" his social environment as he has done in the past, often neglecting changes which others observe. So, although our habitual actions involve guidance by ideals, these ideals are maintained and modified by a complex situation which is the outcome of struggle by a dynamically organized life to maintain itself and to grow within a more or less uncertain and changing environment.

An explanation of any habit must include a view of the organization of the individual's whole body, of the limitations, support and direction offered by the physical environment, and of the way in which the social situation affects the ideal aspect. In the formation and maintenance of a habit these three distinguishable aspects act together in such an intimate way that it is misleading to separate them, except as a device for noting the variety of changes involved. The definition of habit must take account, therefore, of the changing individual and also of those changes in the surroundings which concern him. If, in the midst of all this change, some degree of stability can be discerned, it will be seen as *stability in the situation*. "The situation" includes, of course, the individual and his environment, physical and social, as phases which are inseparable in the explanation of action. The habit is not just *in him*. If a habit is "produced" in a social environment, it will also be "maintained" in a social environment. A habit does not maintain itself.

If a habit seems persistent, then, the student of educational psychology has a problem of measuring the degree of similarity in the physical equipment employed, in the social influences which bear upon the habit, in the organization of the whole body, as well as determining the permanence of the muscular and neural organization immediately concerned. From the standpoint of the teacher,

habit formation involves, therefore, an attempt to modify many of these factors that they may continue to work together in the maintenance of a desirable degree of stability. "Forming" a habit involves something more than so much "practice," with the momentary attainment of a minimum achievement score. And this "something more" can scarcely be described as "overlearning." The educator's duty extends beyond requiring practice to the creation of ideals, and, perhaps, even to the modification of the social environment so that desirable habits may have some chance beyond the confines of the school.

In the face of the common usage of habit as an *explanation* of persistency, it seems a bit brash to assert that the instructor in educational psychology can make it over into a starting post in search of a wide variety of conditioning factors. But, perhaps, the psychologist is not much more seriously handicapped than the physicist who must make clear to his students that "gravitation" is not a peculiar "force" with a "power" of its own. Students of education may yet escape from the notion that "a habit maintains itself," if their instructors place sufficient emphasis upon the interplay of environmental influences with all the activities of the organism in forming and maintaining the particular action labeled "habitual." "Habit" may yet become a stimulus for a type of research which endeavors to face a psychological situation *as a whole.*

At the same time, the student must remember always that the dynamic force which arises out of the organization of a living being as he performs a particular act does make him "take" his environment in a different way than if the act had never occurred. Consequently, it is not an easy matter for us to accept a redefinition of habit, until we have acquired "a special sensitiveness" in that direction; and, then, we already have a new attitude toward habit,—it has taken on new meaning.

GENERALIZATION OF BRIGHT AND DULL CHILDREN,[1] A COMPARATIVE STUDY WITH SPECIAL REFERENCE TO SPELLING

HERBERT A. CARROLL

College of Education, University of Minnesota

The aim of the present study is to uncover through the medium of a school subject (spelling) the comparative ability of bright and dull children to generalize. The approach is made through an analysis and classification of the errors which resulted from the attempts of children to spell unfamiliar words.

Gathering the Raw Data

The subjects from whom the raw data were gathered are fourth and fifth graders in Public School 210, Brooklyn, New York. The causes of misspellings by children in the lower grades are difficult to determine, such a factor as handwriting, for example, often resulting in the appearance of errors even though the child actually knows how to spell the words. With respect to the higher grades, it becomes more difficult to present a novel situation, because of the length and varied nature of previous training. Grades IV and V, then, were thought to be the most satisfactory.

In Grades IV, both A and B (A representing the less advanced half) were chosen. In Grade V it was necessary to omit A, since it was made up of but two sections; hence it offered neither a preeminently dull nor a preeminently bright group. B, however, presented no such difficulties. In IVA there were five sections, in IVB four sections, and in VB four sections. In each of these three grades the brightest and the dullest sections were chosen. These classes had been very carefully differentiated on the basis of intelligence, achievement, and teachers' judgments.

Unfortunately, it was quite impossible for the writer to give Stanford-Binets to the two hundred children serving as subjects in this experiment. Therefore it became necessary to accept the IQ's found on group tests administered at the school. In order to get a more exact measure than would be represented by the score on a

[1] The writer is deeply indebted to the Institute of School Experimentation, Teachers College, Columbia, for the privilege of making this study as a part of the research program of Section D, under the immediate direction of Professor Arthur I. Gates, and the general supervision of the director, Professor Otis W. Caldwell.

single test, the last two IQ's found for the child were averaged and the result set down as the index to the individual's mental ability. (See Table I for details with respect to the composition of the two groups.)

TABLE I.—COMPOSITION OF THE EXPERIMENTAL GROUPS

Class	*N*	Mean IQ	σ	Range in IQ
IVA				
Bright	36	126.4	8.8	101–148
Dull	36	88.5	9.8	66–109
IVB				
Bright	35	110.5	11.3	90–148
Dull	25	89.5	13.7	59–122
VB				
Bright	29	140.6	12.3	120–168
Dull	39	96.8	11.2	66–131
Total				
Bright	100	124.9 ± 1.09	16.3	90–168
Dull	100	92.0 ± .87	12.9	59–131

THE WORD LIST

The words chosen for the purposes of this experiment were taken from the list made by Gates. Gates consulted several compilations of the most frequently used words, such as the Ayres-Buckingham, Horn Basic Vocabulary, Gates Primary, and Thorndike lists. Having made a composite from these, he turned next to the most popular modern spellers in order to discover the grade placement of such words. The spellers used were the following:

1. Horn-Ashbaugh Fundamentals of Spelling
2. Lippincott's New Horn-Ashbaugh Speller
3. The New York City List
4. Jones' Complete Course in Spelling
5. The McCall Speller
6. Smith-Bagley Master Speller
7. Breed-French Speller

He now had a list of words for all the grades, composed of those most often used by children and placed in those grades where, according to the composite opinion of authors of popular spellers, they belonged.

In order to present to the children a situation which was at least comparatively novel, words were given one year in advance of their present grade status; that is, VA words were given to IVA, VB to IVB, VIB to VB. A factor other than novelty was also considered. To

get enough misspellings for experimental purposes, it was necessary to give words considerably in advance of the child's grade placement.

Three hundred eighteen words were dictated to each group. The lists were so broken up that when one teacher had dictated a set of fifty words to a bright group, the same teacher dictated the same set of fifty to a dull group. This equated the teacher factor in so far as presentation was concerned. The teacher first pronounced the word, then used it in a sentence, and then pronounced it again, after which it was written by the children.

Procedure in Analyzing Misspellings

There are a great many classifications possible for a study of this kind. The writer finally decided to use the following:

- Per cent of errors of addition
  - Single letters
    - Doubling
    - Others
  - Groups
- Per cent of errors of omission
  - Single letters
    - Doubling
    - Others
  - Groups
- Per cent of errors of substitution
  - Single letters
  - Groups
  - Entire word
    - Actual word
    - irrelevant
  - Transposition

This classification is a rather rigid one, but it covers every error which children might make although, as has been said before, it would have been possible to have grouped such errors in a different fashion. Before the work of classification had proceeded very far, it became obvious that certain definite and in many cases arbitrary criteria would have to be set up if those analyzing the words were to be consistent both with themselves and with one another. These criteria in every case grew out of actual problems rising in connection with the work of analysis. They cover every instance wherein doubt has arisen in the minds of the analyzers as to just where to place the error. Following are the criteria which were rigidly followed by the three people[1] who analyzed the 32,000 spelling errors appearing in the raw material.

[1] The entire analysis was done by the author with the assistance of Miss Harryette Creasey, a graduate student in English at Columbia University, and Miss Elizabeth Price, a graduate student in Psychology at Columbia University.

Rules Governing Analysis of Misspelled Words

1. When two incorrect substitutions of letters appear in succession, the error is analyzed as a single one and placed under *groups*.

2. When two letters such as *uy* for *ye* in *rye* are substituted, the error is analyzed as: adding of *u* and omission of *e*.

3. A word is put into the column headed *substitution of actual word* only when that actual word is very unlike the one in question. Otherwise it is analyzed, even though it is an actual word.

4. A word is put into the column headed *substitution of an entirely irrelevant word* only when it has little or no structural relationship with the word in question. Otherwise that group of letters which is irrelevant is underlined, and the word placed in the column headed *substitution of a group of letters*.

5. The omission of two or more letters in succession is considered as the omission of a group of letters and looked upon as a single error.

6. Letters are said to be transposed only when they appear in succession as in *laem* for *lame*. Transposition of syllables is not considered as an error to be included in this class.

7. When a letter is added to or omitted after a transposed set of letters, it is said to be added to or omitted after the second letter as it appears in the correct spelling of the word. *Example:* fa*ri*ely. *ri* is transposed. *e* is a vowel added to the consonant *r*.

8. When a single letter is substituted for a group of letters or a group of letters for a single letter, the error is considered as a single one and placed under *groups*.

9. A maximum of three errors in a single word will be analyzed.

10. No specific error will be entered more than once.

11. Each specific error will be entered separately.

Results: Number and Kinds of Errors

Upon correcting the 62,431 words spelled by the children, it was found, as would be expected, that the dull had met with considerably more difficulty than the bright. One hundred dull children spelling approximately the same number of words attempted by a group of bright children of equal size made over three times as many errors; or, putting it in a different way, where the bright misspelled but 18.9 per cent of the words which were one year in advance of their present grade status, the dull misspelled 59.8 per cent. (See Table II.)

There are other data of interest in Table II which might be considered briefly before the analysis of the kinds of errors is approached. It will be seen from the table that the dull were much more likely to be absent than the bright. The causes for such absences were not determined. A factor of greater interest in the present study is the matter of not attempting the spelling of a word. The dull were nearly ten times as dilatory in this respect. The bright, in the face of the difficult

TABLE II.—A COMPARATIVE ANALYSIS OF NUMBER AND KINDS OF ERRORS[1]

			Bright	Dull	Difference
Number of subjects	.....	.....	100	100	
Mean IQ	.....	.....	124.9 ± 1.09	92.0 ± .87	−32.9 ± 1.40
σ	.....	.....	16.3	12.9	
Words omitted	Child absent	.....	150	671	+521
	Spelling not attempted	.....	45	423	+378
Total words spelled	Number	.....	31,675	30,756	−919
	Per cent correct	.....	81.1 ± .148	40.2 ± .189	−40.9 ± .240
	Per cent incorrect	.....	18.9 ± .148	59.8 ± .189	+40.9 ± .240
Per cent of errors of addition	Single letters	Doubling	3.0 ± .133	1.5 ± .052	− 1.5 ± .142
		Others	14.6 ± .276	13.1 ± .143	− 1.5 ± .311
	Groups	.....	1.0 ± .077	1.4 ± .049	0.4 ± .091
Per cent of errors of omission	Single letters	Doubling	6.1 ± .187	3.4 ± .077	− 2.7 ± .202
		Others	26.6 ± .368	20.7 ± .172	− 5.9 ± .406
	Groups	.....	2.6 ± .124	5.2 ± .095	+ 2.6 ± .156
Per cent of errors of substitution	Single letters	.....	24.6 ± .337	22.3 ± .178	− 2.3 ± .381
	Groups	.....	14.2 ± .274	21.0 ± .174	+ 6.8 ± .325
	Entire word	Actual word	0.6 ± .006	1.5 ± .052	+ 0.9 ± .052
		Irrelevant	0.8 ± .069	7.3 ± .111	+ 6.5 ± .131
	Transposition	.....	5.8 ± .183	3.3 ± .076	− 2.5 ± .198
Total number of errors	.....	.....	7,416	24,882	+17,466

[1] Throughout this report a minus sign will be used to indicate that the dull rate less than the bright; a plus sign, more.

situation resulting from being asked to spell an unfamiliar word, are much more likely than the dull to try.

The differences in kinds of errors made are important. It will be kept in mind that in arriving at the conclusions presented in Table II over 32,000 errors were classified into eleven groups. This represents a considerable sampling. A study of the table brings to our attention the fact that in each of the classifications a significant difference between bright and dull is found, accepting as statistically significant a Critical Ratio of three or more. In some types of errors the bright exceed the dull; in others the dull exceed the bright. The bright make more errors in the following instances:

1. Addition of a single letter through doubling
2. Addition of single letters which do not involve doubling
3. Omission of the second of a double

4. Omission of a single letter not one of a pair
5. Substitution of one letter for another
6. Transposing letters

The dull excel the bright in the following types or errors:

1. Addition of a group of letters
2. Omission of a group of letters
3. Substitution of a group of letters
4. Substitution of an entire actual word
5. Substitution of a wholly irrelevant word

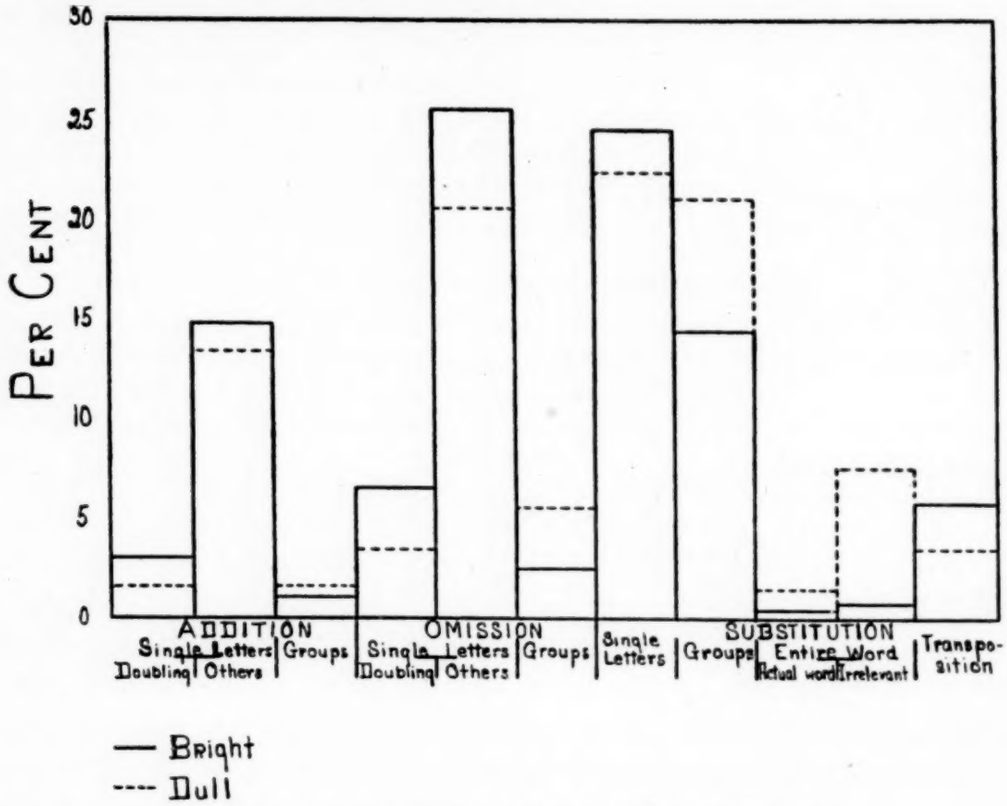

FIG. 1.—A comparative analysis of kinds of errors.

A rather striking fact emerges from this summary of the differences; namely, that the dull are likely to make a more widespread error; that is, they are likely to add, omit, or substitute a group of letters. The bright, on the other hand, appear to have greater difficulty, speaking always on the basis of comparative per cents, with single letters. (See Fig. 1.)

In terms of Critical Ratio the greatest difference between the bright and dull appears in the irrelevant column, where it amounts to 6.5 per cent with a critical ratio of 49.6. It is obvious with such high reliability that a very real difference exists between bright and dull

in the number of wholly irrelevant spellings made. The second highest Critical Ratio, 20.9, appears in that class of errors designated as the substitution of a group of letters. Included in this class are those groups of letters which, though they are wholly irrelevant, have not so vitally affected the word as to make it unrecognizable, thus making it necessary to place it in the wholly irrelevant column. The next greatest difference appears in the substitution of an actual word for the word dictated. Again the dull appear to be quite at sea. If the word *picture* is given, it is not at all surprising for the child to react by writing the word *boat.*

While the three largest differences appear as errors of substitution, the fourth most reliable difference, with a Critical Ratio of 16.7, is in the omission of an entire group of letters. Again the dull seem to be unable to catch the rhythm of the word; as a result, they omit a syllable or some other group of letters. The greater per cent of errors in the next six groups appears on the bright side of the ledger. It will not be necessary to take up each of these in turn, but it is rather interesting to note that the bright seem to have considerable difficulty with doubles, and tend to a greater degree of errors of transposition.

Comparison between Ten Brightest and Ten Dullest from Each Grade

In order to deal with the more theoretical problem of the differences between the bright and dull, as measured by intelligence tests rather than as grouped by the intelligence quotient plus other criteria, the ten highest and the ten lowest IQ's were chosen from each of the three grades—IVA, IVB and VB.

This re-classification also serves as a check on a possible flaw in the division of bright and dull in the main groups, where the number of each type was unequal in two of the three grades, IVB and VB, though the totals were exactly equated at one hundred. This meant that identical opportunity to make certain types of errors was not furnished the two groups. The difference, however, was very slight. But even this objection is eliminated in the present grouping where number of children and number and kinds of words are identically equated.

With the greater difference between the mean IQ's of the two groups—for in this case the mean of the bright group is 137.8 and of the dull 78.7—it would be expected that there would be a greater divergence in both number and kinds of errors. This is exactly what happens, and bears out the theory that the kinds of spelling errors

made by children depend to some extent upon their intelligence. (See Table III.)

TABLE III.—A COMPARATIVE ANALYSIS OF NUMBER AND KINDS OF ERRORS MADE BY THE TEN BRIGHTEST AND TEN DULLEST CHILDREN IN EACH GRADE

			Bright	Dull	Difference
Number of subjects	.....	.....	30	30	
Mean IQ	.....	.....	137.8 ± 1.785	78.7 ± .799	−59.1 ± 1.958
σ	.....	.....	14.5	6.5	
Words omitted	Child absent	.....	0	121	+121
	Spelling not attempted	.....	13	227	+214
Total words spelled	Number	.....	9547	9333	−214
	Per cent correct	.....	84.4 ± .250	34.0 ± .330	−50.4 ± .414
	Per cent incorrect	.....	15.6 ± .250	66.0 ± .330	+50.4 ± .414
Per cent of errors of addition	Single letters	Doubling	3.5 ± .298	1.3 ± .084	− 2.2 ± .310
		Others	13.1 ± .548	13.4 ± .252	+ 0.3 ± .603
	Groups	.....	0.8 ± .144	1.9 ± .101	+ 1.1 ± .176
Per cent of errors of omission	Single letters	Doubling	6.7 ± .406	2.6 ± .117	− 4.1 ± .423
		Others	26.8 ± .719	20.3 ± .297	− 6.5 ± .778
	Groups	.....	2.9 ± .272	5.8 ± .173	+ 2.9 ± .322
Per cent of errors of substitution	Single letters	.....	25.0 ± .703	21.9 ± .305	− 3.1 ± .766
	Groups	.....	13.1 ± .548	21.5 ± .303	+ 8.4 ± .626
	Entire word	Actual word	0.8 ± .144	1.4 ± .087	+ 0.6 ± .168
		Irrelevant	1.1 ± .169	6.6 ± .183	+ 5.5 ± .249
	Transposition	.....	6.1 ± .388	3.2 ± .130	− 2.9 ± .409
Total number of errors	.....	.....	1726	8321	+6595

In so far as number of words misspelled is concerned, it will be noticed that the difference between the total number of the thirty brightest children and the thirty dullest children is 10 per cent greater than was the difference between the two main groups where a smaller spread between the mean IQ's existed. That, however, would be expected. What is of greater significance is the fact that in eight of the eleven classes of errors the difference between bright and dull is intensified. Only in the addition of a single letter, which is not a double, the substitution of an entire actual word, and the substitution of a wholly irrelevant word, is there a decrease of the span between the two per cents. Close observation of Table III will reveal the fact that the differences are increased from .03 per cent to 1.6 per cent.

This is not very great, to be sure, but it is large enough to be significant. It is rather difficult to understand the unexpected change in the "addition of a single letter." Not only is there an alteration in this case of 1.8 per cent, but the sign is reversed, the thirty dullest children making more errors than the thirty brightest ones, while when the two main groups were being considered, the bright group made more errors than the dull. These differences in the other two classes mentioned are not so great and can be accounted for in part by the greater probable error.

Interpretation

The results just stated showed that in certain kinds of errors bright children excelled the dull, while in others dull children excelled the bright. The interpretation of such differences must necessarily be somewhat speculative.

The bright, it was found, consistently have much greater difficulty with single letters. (It will be kept in mind that all these comparisons are made on the basis of per cents. In no type of error did the bright exceed the dull in raw score.) This difficulty with single letters may be due to the fact that they offer greater possibilities for phonetic renderings than do groups of letters. That is, a child capable of a high degree of generalizing is more likely to spell *cattle, k-a-t-l-e*, than he is to spell it *s-z-a-t-b-m-l-e*. In the first misspelling, the child has made two single letter errors, each of which is phonetic. Probably similar phonetic generalizations account, then, for the preponderance of single letter errors by the bright.

Single letter errors accounted for six of the eleven types in the classification used. The remaining five were mistakes involving groups, in each of which the dull excelled the bright. The explanation seems to be the one just offered relative to single letter errors; in this case, the dull, failing to make phonetic renditions of the words given them, add, omit, or substitute groups of letters which in the great majority of cases bear little if any association to the sound qualities of the word. Thus to spell *table, t-x-m-b-l-e*, shows a lack of ability in translating the sound of long *a*.

It will not be necessary to give a detailed interpretation of the differences found in each of the eleven types of errors utilized. They have been treated in general in the above discussion. Some of the outstanding variations, however, may well be considered briefly. This investigation shows that the bright child is twice as likely as the

dull to make the error of doubling a letter. For example, he may spell *advise, a-d-d-v-i-s-e.* Excellent phonetic transfer! He is given two sounds to translate into letters, and he errs in spelling the sound *ad, a-d-d.*

The same situation exists in a second instance in which he makes a greater percentage of errors than the dull; namely, that of omitting the second of a double. When he spells *rabbit, r-a-b-i-t,* it is obvious that he is not violating the principles of good phonetic generalization procedure. Though he has omitted the second *b,* the word sounds satisfactory. If the child had gone further and broken the word up into its two pronunciation units, perhaps the mistake would not have been made. Evidently he failed to make this type of motor transfer.

A large difference was found between bright and dull in the number of errors of transposition made, the bright exceeding the dull by a considerable amount. Perhaps here, as in the preceding instances, the explanation lies in the fact that the bright generalize phonetically more than the dull. Such a word as *scheme,* for example, might easily be rendered as *s-c-h-e-e-m.* The child has transposed the *m* and the *e,* but he has given a phonetic spelling, nevertheless. To check up on this supposition, the investigator classified into phonetic and non-phonetic groups three hundred transposition errors chosen at random from the bright list. One hundred ninety-one proved to be phonetic and one hundred nine non-phonetic. On classifying the same number of dull errors of this type, it was found that one hundred sixty-four were phonetic and one hundred thirty-six non-phonetic.

This represents a difference in raw scores of twenty-seven points—not quite so great as was expected. Evidently errors of transposition have a tendency to be phonetic when made by either bright or dull, although the bright have an advantage in this respect.

While the dull group excelled the bright in none of the single letter errors, they did so in every one of the five group errors. Noticeable among these was that of substituting a wholly irrelevant spelling, such as *rszilm,* for the word *defeated.* In this kind of misspelling, the dull had a per cent of error more than nine times as great as the bright. The psychological explanation appears to lie in the inability of the dull to adjust themselves to a new situation, meeting it by drawing more or less heavily upon past experiences. Such wholly irrelevant spellings are illogical; the child has failed to recognize in the word a single element identical with some element in his past spelling experience. The visual appearance of the word is markedly, if not wholly,

different from the one dictated; the pronunciation units have been completely lost; and no phonetic transfer whatever has been made. This type of error represents the best example in spelling of an almost total lack of ability to generalize. It is important to remember that it is an error which the dull child is far more likely to make than the bright. The bright child does not fall into such a widespread and illogical error because he is too much on the alert for transfer possibilities.

The remaining four types of errors, involving groups, in which the dull excel the bright, have a somewhat similar psychological background. A wide difference, for example, exists between the two classes of children in the matter of substituting one group of letters for another. Included in this class of errors is the substitution of an irrelevant group of letters—a group, however, that is not large enough to cause the word to be placed in the wholly irrelevant word column. Again the dull show an inability to generalize. For example, if a child spells *monument*, *m-o-n-s-l-i-g-h-s*, he has substituted a group of irrelevant letters for that part of the word which should have been spelled *ument*. There appears to be no logic behind such a substitution. Perhaps some slight signs of visual transfer might be discovered with respect to the length of the word, but the pronunciation units have been entirely lost, and there is not a bi it of phonetic transfer. In the light of these facts, it becomes significant that the bright are much less likely than the dull to make this kind of error.

Summary

1. Marked differences in degree were found to exist between bright and dull in kind as well as in number of errors, the bright showing a greater tendency to err with respect to single letters, and the dull with groups of letters.

2. An analysis of the errors made by the ten brightest and the ten dullest children in each grade results in an even greater gap between the per cents of types of errors made by bright and dull.

3. It appears that phonetic generalization is the dominating factor in the psychology of the differences in degree in the kinds of spelling errors made by bright and dull. The bright child, possessing a high degree of ability in translating sounds into letters, makes those mistakes which would quite naturally result from the varied phonetic qualities of the English language; while the dull child, less capable of adapting himself to a new situation, makes those errors which have little if any phonetic foundation.

PHONIC VERSUS LOOK-AND-SAY TRAINING IN BEGINNING READING

RAYMOND M. MOSHER

University of Idaho

AND

SIDNEY M. NEWHALL

Yale University

INTRODUCTION

The literature on the relative merits of the look-and-say and the phonic methods has been increasing. Gates has indicated a suspicion that the phonic method has failed to establish essential habits in reading. He writes: "Grave difficulties attend this method of depending on supplementary training to develop various desirable habits in reading." And, "In other studies, similar limitations have been found for certain types of phonetic drill."[1]

Sexton and Herron in an experiment carried on in the Newark, New Jersey, schools conclude that the value of phonics is doubtful in the first five months of the first grade. They say: "Although the outcome of the experiment tends to favor some phonic instruction beginning in the second half of the first year, just what type of phonics should be taught and to what extent, are still problems for experimentation."[2]

Mosher has shown that children may acquire a large vocabulary and have necessary equipment for meeting new reading situations without the use of phonic training.[3]

Jagger opposes the phonic method on the grounds that it is as synthetic in character as the alphabet method.[4]

Wheat advocates the teaching of a method of attack on new printed words in which the words are presented in meaningful sentences

---

[1] "The Improvement of Reading." Pp. 26, 27. Macmillan, 1927. *Cf.* also, "New Methods in Primary Reading." P. 18. Bureau of Publications, Teachers College, Columbia University, 1928.

[2] *Elementary School Journal*, Vol. XXVIII, 1928, pp. 690–701.

[3] Some Results of Teaching Beginners by the Look-and-say Method. *Journal Educational Psychology*, Vol. XIX, No. 3, pp. 185–193.

[4] "The Sentence Method of Teaching Reading." Grant Educational Co., Ltd., London, England, 1929.

and the pupil is "led to check the result of his examination of the new printed word against its familiar spoken counterpart."[1]

Definition of Terms.—As defined for this experiment, the look-and-say method of teaching reading seeks to attach meaning to the written symbols without analytical means for aiding in the correct response. This means that the emphasis is placed on reading a word or group of words as a unit printed symbol and thinking the thing for which the symbol stands. Discussion of the phonic qualities of vowels, consonants or phonograms was not a part of the instructional procedure. The phonic method differed from the look-and-say method by providing systematic drill on the phonetic qualities of words. For example, likenesses in words were introduced by such devices as rhymes or jingles. The aural qualities were then associated with their corresponding written symbols. Drills followed to aid the pupils in the recognition of words belonging to certain families.[2]

Formal instruction in phonics was limited to a period of fifteen minutes per day during the first year. This time allowance was added to that provided for reading in the daily class room schedule.

In both procedures, the initial stages were the same, in that the teachers urged the children to develop reading themes concerning home, school and every-day experiences which were written on the blackboard or chart and read by the children. The vocabulary was that suggested by the children, with reasonable guidance by the teacher. These themes constituted the reading materials for all the pupils (phonic and look-and-say) during the first six weeks of the first grade on the assumption that the materials would develop a favorable mind-set for reading and that essential vocabulary preparatory to book reading would be introduced.[3] When the six weeks of introductory work had been completed, primers were used. The transition was apparently easy to effect, and at this period the phonic group began the systematic training mentioned above.

Experimental Problem

The purpose of this experiment is the comparison of reading habits resulting from two methods of teaching beginning reading, the

[1] Examination Analysis versus Phonetic Analysis in Primary Reading. *Elementary School Journal*, Vol. XXIX, No. 4, pp. 256–266.

[2] The elements listed in the Twenty-Fourth Year Book, National Society for Study of Education, pp. 88–89, were used by teachers as a guiding list of minimum essentials in phonic training.

[3] *Cf.* Mosher: *Op. cit.*, p. 187*ff*.

phonic and the look-and-say. More specifically, such phases as frequency of eye-fixations, speed on easy and difficult material, and comprehension were studied and quantitative measures secured in the effort to determine differences characteristic of one or the other of the groups.

Reading is admittedly a complex matter. To secure measures of all the phases of it, while such might be very desirable, is unfortunately not easy or necessarily possible. It is conceivable, too, that one method of measurement might favor one group while another might not, and that therefore a poor selection of measures might lead to an invalid conclusion. This we tried to avoid by using fundamentally important measures and applying them in such a way as would resemble ordinary procedures of the reading class as much as practicable. The tests for eye-fixations were introduced as a cooperative game. In view of the fact that also there was notable eagerness and willingness to participate on the part of the subjects, it seems reasonable to believe that the data of the present investigation may reveal possible *essential* differences in results of the two reading methods.

The subjects were from the three elementary schools connected with the State Normal School, New Haven, Connecticut. When these children entered the first grade in September, 1926, they were tested by the use of group intelligence tests and sectioned so as to have superior, average, and below average pupils in every first grade class room. There were two first grade class rooms in each of two of the schools and three in the other school. In all three buildings the look-and-say method of teaching reading was employed in one of the first grade class rooms while the phonic method was used in the other room (or rooms). At the end of the year the pupils were promoted to the second grade rooms and the same general reading policy continued until the end of that year. Since the phonic and look-and-say groups may thus be regarded as approximately random and equated as to intelligence and since there is no reason to believe that the quality of the teaching in the two groups was essentially different, the comparison of elements of reading proficiency at the end of the second year seems legitimate. The data presented here were secured in April and May, 1928, and include fifty look-and-say and seventy-three phonic children from whom complete test records were available. Many more were tested on certain aspects of the problem whose records

were incomplete in one, two or more items; they were eliminated in the final tabulations.[1]

Procedure

The tests which were devised to measure the reading habits under consideration in this study were applied under as univocal conditions as practicable. A relatively quiet and otherwise unused room was employed, and each subject was admitted separately. The children were tested in a random order so that the examiners never knew whether they were testing a phonic or a look-and-say child. This fact would seem to eliminate the possibility of partiality to either group. A description of the two testing procedures follows:

Procedure A—Tests of Eye-habits.—The subject was seated in position before an instrument which held the reading material and supported his head. The instrument was devised for this experiment to make possible the estimation of the number of eye-fixations per line.[2]

The "easy" selection, read orally, consisted of four sentences equal in length, the words equivalent in difficulty to ordinary primer material. The choice of words was from the first two hundred in Thorndike's word list. The sentences were typed on a card, the primer-type typewriter being used, in such a way that they appeared as a full paragraph. The "difficult" selection was equal in difficulty to typical first reader material. The typing and general arrangement on the page were similar to that of the "easy" selection.

Procedure B—Oral and Silent Reading.—This procedure was informal in that the child faced no apparatus but read the selections, one easy and two difficult, from a book held in his hand while he sat or stood as he chose.[3] The oral reading yielded data on word recogni-

---

[1] The authors are indebted to Mrs. Frances Furbish, Mr. Herbert Clish and Mr. Wilfred Moody, Principals of the Elementary Schools where the experiment was carried on; also to Miss Lucy Stone and especially to Miss Iva Mercer for assistance in administering the tests.

[2] The instrument facilitates the observation of eye-movements or pauses chiefly by presenting the experimenter with an enlarged, illuminated image of the subject's eye. Apparatus and technique are fully described by one of the writers (S. M. Newhall) in another article in the *American Journal of Psychology*, Vol. XL, 1928, p. 628*ff*.

[3] "The Wise Old Cat," pp. 55–58 and "The Policeman," pp. 119–120 from "The Child Story First Reader," Lyons and Carnaham, 1927, and "The Boy and Goat," Free and Freadwell Primer, Row Peterson and Co., 1910. The first two are referred to as difficult and the third, easy.

tion, which will be reported in a subsequent article, as well as a time estimate on each of the three selections. It will be noted that this procedure closely approximates normal conditions of a reading class in which a minimum of restriction is placed upon the child, in contrast with Procedure A. Immediately following the oral reading, the child was given silent reading comprehension tests on each of the three stories. Estimates of speed in completing the tests were obtained also. Small card-board squares upon which were printed *yes* and *no* were given the pupil and he placed the appropriate answers alongside the questions in accordance with his opinion.

Results

Tables I to VI give the averages and standard deviations for the two groups in the various tests which were applied. It will be noted that these measures are generally in rather close agreement, a fact which seems to indicate on its face that there is little advantage of the phonic over the look-and-say group.

This indication is supported by the lack of statistical significance of the differences between the two groups in the various tests. For example, Table I presents the means and σ's for the eye-fixations per

Table I.—Number of Eye-fixations per Line for Easy Selection

	Look-and-say	Phonic
Number	50	73
Mean	7.24	6.81
σ	3.31	3.38

Table II.—Number of Eye-fixations per Line for Difficult Selection

	Look-and-say	Phonic
Number	50	73
Mean	8.96	8.36
σ	3.65	3.77

Table III.—Number of Seconds to Read Easy Selection

	Look-and-say	Phonic
Number	50	73
Mean	35	31.58
σ	27.4	24.2

Table IV.—Number of Seconds to Read Difficult Selection

	Look-and-say	Phonic
Number	50	73
Mean	46.6	43.91
σ	29.2	29.43

TABLE V.—NUMBER OF SECONDS ON SILENT READING

	A, difficult (Wise Old Cat)		B, easy (Boy and Goat)		C, difficult (Policeman)	
	Look-and-say	Phonic	Look-and-say	Phonic	Look-and-say	Phonic
Number	50	73	50	73	50	73
Mean	122.8	118.77	103.2	102.91	107.6	124.8
σ	62.45	49.8	39.69	39.36	48.6	56.84

TABLE VI.—PER CENT OF ACCURACY ON COMPREHENSION TESTS

	A, difficult (Wise Old Cat)		B, easy (Boy and Goat)		C, difficult (Policeman)	
	Look-and-say	Phonic	Look-and-say	Phonic	Look-and-say	Phonic
Number	50	73	50	73	50	73
Mean	77.4	80.75	90	94.17	84.2	78.56
σ	18.6	17	12	17.7	16.9	18.9

line. The standard error of the difference between the two means may be shown by an application of the usual formula:

$$\sigma_{\text{diff.}} = \sqrt{\sigma^2(\text{av. } 1) + \sigma^2(\text{av. } 2)}$$

In this particular example we find that $\sigma_{\text{diff.}} = .60$. Dividing the difference between the means, *viz.* .43 by .60, we find that $\frac{\text{diff.}}{\sigma_{\text{diff.}}}$ = only .72. The differences in means between the two groups in the other tests seem to be about equally lacking in significance. In no case is $\frac{\text{diff.}}{\sigma_{\text{diff.}}}$ as large as 2. These values are shown in Table VII.

Though the observed differences are thus all comparatively small it must be noted that in eight out of the ten cases the differences are in the direction which favors the phonic means. On the other hand, the two largest differences of all are those which favor the look-and-say means. On the whole, the quite frequently made assumption that phonic training leads to substantially more accurate and independent reading appears to be negated.

TABLE VII.—STANDARD ERRORS OF THE DIFFERENCES BETWEEN PHONIC AND LOOK-AND-SAY MEANS
(Calculated from Tables I–VI)

	Diff.	$\sigma_{Diff.}$	$\frac{Diff.}{\sigma_{Diff.}}$
Table I	.43	.60	.72
Table II	.60	.67	.90
Table III	3.42	4.78	.72
Table IV	2.69	2.74	.98
Table V*a*	4.03	10.56	.38
b	.29	7.26	.04
c	17.20	9.56	1.80
Table VI*a*	3.35	3.30	1.02
b	4.17	2.68	1.56
c	5.64	3.25	1.74

SUMMARY

The indications of the data appear quite consistent. Since the data are comparatively limited, however, our interpretations of them must be regarded tentative in respect of general application.

1. The measures of speed, fixation pauses, and comprehension seem to show, both individually and collectively, no essential difference in results of the two training methods compared, at least at the end of the second year.

2. If these measures comprise a valid index of reading proficiency then the supposed advantages accruing to phonic training are negated. Look-and-say children under the experimental conditions maintained in the schools are able to read approximately as quickly and effectively as phonic children.

3. In general our results suggest that added time devoted to phonics would not appreciably increase reading skill, and that therefore phonic training is not especially to be recommended as a device for that purpose.

A COMPARISON OF DIFFICULTY AND ACCURACY IN THE LEARNING OF BRIGHT AND DULL CHILDREN IN A MOTOR-MEMORY TASK

FRANK T. WILSON

State Teachers College, Buffalo, N. Y.

In a previous study[1] the learning of bright and dull children was compared in the task of placing the blocks of the Goddard formboard into their proper places with the eyes blindfolded after two trials with sight. It appeared from that study that for groups of children nine and twelve years of age, half of each group at a level of seventy to eighty IQ and half at one hundred ten to one hundred twenty, the average difference of thirty IQ points gave advantage, as measured by amount of work, to the brighter children, and that the difference of three years favored the older groups. The difference in chronological ages, however, seemed to give the greater advantage, rather than the difference in IQ.

This report gives the result of further study of the groups in this task in regard to difficulty of pieces and from the records of errors committed. Data for the groups are given in Table I, eight additional cases having been included in the later study.

TABLE I.—DESCRIPTION OF SUBJECTS

	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Number of cases	15	19	19	15
Average CA	9-1	9-2	12.2	12-2
CA range	8-11 to	9-4	12-0 to	12-5
Average MA	7-9	10-5	10-4	13-11
Average IQ	85.1	114.0	85.0	115.0

The following explanation of errors was adhered to: " . . . each attempt to place a piece in a wrong recess being considered an error, provided the piece touched either the bottom or the inner edge of the recess; but continued efforts with the same piece at the same place were not counted as additional errors. If, however, an attempt was made to place a piece in a second recess and then effort was resumed at the first place, the resumed attempt was counted as an additional

[1] Wilson, Frank Thompson: "Learning of Bright and Dull Children." Bureau of Publications, Teachers College, N. Y.

error. Subjects were not told that errors were being counted and no report on them was made to the children."[1]

Table II gives the average number of errors by groups and sex.

TABLE II.—AVERAGE NUMBER OF ERRORS BY GROUPS AND SEX

	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Boys	101	102	87	74
Girls	155	117	110	77
Boys and girls	126	111	97	75

In addition to the advantages of age and IQ noted above there is a superiority of boys over girls at both ages and at both IQ levels.

Three further questions in regard to the character of the learning of these groups were studied and will be discussed.

1. Are there differences in the difficulty which the various blocks presented to the four groups? The answer to this question was determined by totaling the number of errors made with each piece by each of the four groups. The counts were then ranked. Table III gives the ranking of the pieces from low to high, first for the entire group, and then the lowest three and the highest three for each group.

TABLE III.—ORDER OF DIFFICULTY FROM LOW TO HIGH ACCORDING TO NUMBER OF ERRORS

	Total	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Half circle	1	3	1	1	2
Cross	2	2	..	2	1
Star	3	1	2	3	
Circle	4	..	3		
Lozenge	5	8			
Oblong	6				
Triangle	7	..	..	..	3
Oval	8	..	8	9	10
Square	9	9	9	8	8
Diamond	10	10	10	10	9

It is to be seen that on the whole the same pieces were the easiest for each group and the same ones were the hardest for each group. Why should the half circle, cross and star be the easiest of the ten pieces

[1] *Ibid.*, p. 11.

for young dull children and also for bright older children? Why should the oval, square and diamond be the hardest ones for bright and dull, young and old? No clear answer seems suggested by the data. Whatever accounts for this somewhat remarkable fact it seems safe to observe either that brightness and past experience of these subjects had nothing to do with the relative difficulty of the blocks used, or that past experience in the lives of these subjects gave such training that it was easier on the average for all subjects to place the half circle, cross and star in their proper places and harder to put the oval, square and diamond forms where they belonged. Possibly confusion due to similarity of square and diamond explains their difficulty. If that is so there is still the unexplained fact that those same pieces are hardest for bright as well as for dull. Possibly the oval was confused with the circle and half circle. The rank of the half circle, however, is one, that of the circle four and the oval eight. No satisfactory explanation for these facts comes to the mind of the writer.

2. What differences can be found in the progress of learning as measured by the reduction in the number of errors? Table IV*a* shows the average number of errors for each trial for each of the groups.

TABLE IV*a*.—AVERAGE NUMBER OF ERRORS FOR EACH TRIAL FOR EACH GROUP

	Trials											
	1	2	3	4	5	6	7	8	9	10	11	12
Dull, 9	15	17	14	14	17	12	8	8	5	6	5	5
Bright, 9	11	16	11	16	9	12	8	7	5	6	3	3
Dull, 12	13	11	11	11	6	7	7	7	5	6	6	5
Bright, 12	9	9	10	9	7	6	7	5	6	4	3	2

In the main all groups show steady decrease. Table IV*b* translates the data of IV*a* into per cent for certain parts of the learning record.

TABLE IV*b*.—PER CENT OF ERRORS WHICH FOUR SELECTED PARTS OF THE LEARNING SERIES ARE OF THE WHOLE SERIES

	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Trial 1	12	10	13	13
1, 2	26	25	24	25
1, 2, 3	37	34	36	38
1–6	71	69	63	67

It can scarcely be thought that the differences in the figures of this table represent real differences. Whether one, two, three or half of the trials are considered, the percentages of total errors which these trials show are almost exactly the same for all the groups. Of course the criticism of such figures is that the units involved are of unequal value, probably, in that improvement from nine errors to two is not strictly comparable to an improvement from fifteen to five. For this reason it would seem plausible to suggest that for bright twelve to maintain as high a percentage of improvement as the other groups, when the number of errors for that group is smaller from the first through nearly every trial to the last, is to make a better record.

3. What differences are shown as to the persistency with which errors recur? An answer to this question is attempted by counting the subjects who made errors in seven or more of the twelve trials with any piece. This number, seven, represents more than half of the series of trials for each piece and indicates, in a way, the persistency of recurring difficulty encountered in handling the piece. Evidently if errors were made with a given piece in at least seven trials there must have been some unusual difficulty. Whether or not it is correct to call the error recurring is open to question, as the nature of the errors may have been different in each instance. It is believed, however, that the figures have significance.

Table V*a* gives the percentage of subjects having errors in seven or more trials for any piece.

TABLE V*a*.—PER CENT OF SUBJECTS HAVING ERRORS IN SEVEN OR MORE OF THE TRIALS FOR ANY PIECE

	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Boys	90	50	70	60
Girls	100	55	60	50
Total	93	52	68	51

Differences between boys and girls seem uncertain, but in the totals it seems that the bright subjects do better than the dull—age notwithstanding. For the dull, age may explain an apparent difference. For the bright, there seems to be no difference despite age.

Table V*b* shows the same thing. The figures in this table are slightly different, being the per cents of the pieces of all subjects of each group during the learning series of which there were seven or more trials with errors.

TABLE Vb.—PER CENTS OF PIECES DURING TRIAL SERIES OF WHICH THERE WERE SEVEN OR MORE TRIALS WITH ERRORS

	Dull, 9	Bright, 9	Dull, 12	Bright, 12
Boys	16	15	19	19
Girls	30	17	20	11
Total	23	16	20	15

The total twenty-three in this table means that of all trial series with all the pieces the dull nine year old had errors in twenty-three per cent of the cases. The same facts seem to be shown in this table as in the table of the percentage of subjects making errors in seven or more trials, namely, a better score for the older dull group than for the younger, and better scores for the bright groups than for the dull, there being a difference of but one between the error for the two bright groups. Brightness seems to be much more significant than age in years in this respect.

Conclusions.—In a learning experiment in which subjects, after two visual practices, placed the blocks of the Goddard formboard into their proper places with their eyes blindfolded, it was found that

1. The same three of the ten pieces were hardest and the same three easiest for dull and bright, young and old children.

2. Fewer errors occurred on the part of the brighter and older children and that even with this smaller beginning number of errors the percentage of reduction of errors was as high for them as for those having a larger number of beginning errors.

3. Errors persisted more generally in the efforts of duller children and that a greater number of persisting errors were found in the records of young and dull children.

4. Although amount of work showed only a small superiority of bright and old over dull and young, accuracy, as indicated by errors, showed a considerable superiority for the brighter and older, the superiority of the brighter being especially clear.

FACTORS DETERMINING RELIABILITY OF TEST NORMS

E. F. LINDQUIST

State University of Iowa

The purposes of this article are: (1) To establish empirically the relative importance of certain factors determining the reliability of norms for standardized tests, particularly of high school subjects; (2) to demonstrate, on the basis of objective evidence, that the influence of certain hitherto neglected factors may be such as to render all but a very few of the norms now available for high school tests practically worthless for the evaluation of school achievement on a relative basis; (3) to suggest a tentative criterion of reliability for a test norm; (4) to determine empirically the specific conditions under which this criterion will be satisfied; and (5) to suggest certain changes in the present ways of expressing norms of performance for standardized tests.

The data used to substantiate many of the conclusions here drawn were secured through the testing of over 30,000 high school pupils in 223 Iowa high schools, resulting in more than 53,000 pupil-testings. These data were made available through the Iowa Academic Contest of 1929. This contest, designed to stimulate scholastic endeavor and to serve as an instrument for curriculum revision in the state of Iowa, was conducted in a series of three progressive competitions, only one of which—the school contest—is of immediate concern here. For this school contest there was constructed a forty-five-minute objective achievement examination of each of twelve high school subjects. Ten of these subjects will be found listed in Table I. The other two subjects, Stenography and Typewriting, are omitted from this discussion because of insufficiency of data—the number of schools competing in these subjects being relatively small.

Each school entering the school contest was required, in order to be eligible for competition, to enter and test every student in all classes in the subjects in which the school competed. Any school showing evidence of selection of superior students has been eliminated from the analysis here presented. For the schools involved, therefore, the data herein contained are based on the scores of *all* students taught in the subjects listed.

Each of the examinations was completely objective, in each the types of test items used were very similar to those found in current

standardized tests, and each was carefully constructed on the basis of the usual validation techniques and of preliminary try-outs outside of the state of Iowa. The reliability coefficients, established in the preliminary try-outs, were consistently high, showing a value of .92 or better for more than half of the tests. For the reasons mentioned, therefore, it is believed that the data here presented may be fairly taken as representative of those which might have been secured from similar administrations of any good standardized tests now on the market for the same high school subjects.

Defects of Current Test Norms.—The writer has already stated his conviction that, for the majority of tests now published for high school subjects, the norms provided are so unreliable and ill-defined as to be practically without value for the purpose of evaluating school achievement on a relative basis. Objective evidence will shortly be presented to show the reasonableness of this statement. Among the reasons for the breakdown of test norms are the following: (1) The fallacious practice of playing up mere *size* of the sample used for standardization as of major importance; (2) failure to *define the population* for which the norm is established and to *select* the sample so as to be fairly representative (random) of any meaningful population; (3) failure to recognize that no single population can be defined so as to be of value to all types of schools, but that schools must be classified into comparable groups and norms established independently for each group; (4) failure to take into adequate consideration the variability in *school* achievement relative to that in achievement of individual pupils; and, (5) failure to provide measures of central tendency and of variability of school averages as well as of raw scores.

Magnitude of School Differences.—It will not be possible, within the limits of this brief discussion, to consider the relative importance of all of the factors just mentioned. The one of greatest importance, however, because of its bearing upon each of the others, is perhaps the factor of variability in achievement where the *school*, rather than the pupil, is considered as the unit. It is to this factor, therefore, that the major share of our present consideration will be given.

The fact that large achievement differences *between schools* do exist in amount greater than can be accounted for by chance factors has long been recognized and considered important by authorities on test construction. It is extremely doubtful, however, that the enormous relative magnitude of such differences has been fully appreciated, or that all the resulting implications relating to the problem of sample

selection have been adequately recognized. The major contribution of this article, therefore, will lie in the presentation of objective evidence demonstrating empirically the relative magnitude of school differences. This evidence is presented in Table I:

TABLE I.—VARIABILITIES IN RAW SCORES AND IN SCHOOL AVERAGES COMPARED

Test	(1) Number of pupils	(2) Standard deviation of raw scores (σ_R)	(3) Number of schools	(4) Standard deviation of school averages (σ_A)	(5) $\frac{\sigma_R}{\sigma_A}$
1. English Mechanics—9th	7036	16.60	207	6.33	2.62
2. English Mechanics—10th	6770	18.73	208	8.31	2.25
3. Eng. and Amer. Literature—11th	5323	24.03	188	12.51	1.92
4. Eng. and Amer. Literature—12th	3233	26.26	172	15.81	1.66
5. Ninth Year Algebra	6684	6.81	206	4.84	1.41
6. Plane Geometry—10th	6304	5.36	209	3.12	1.72
7. General Science—9th	4460	20.23	164	10.41	1.94
8. Physics—12th	4261	20.31	193	9.89	2.05
9. American History	5396	19.93	157	9.51	2.09
10. World History	4202	22.41	171	12.09	1.85
				Average =	1.95

The table should be read as follows, using the test in English and American Literature for Grade XII as an illustration.

This test was given to 3233 pupils (column 1) in 172 different high schools (column 3). The standard deviation of raw scores (column 2) for the total 3233 pupils was 26.26 score units. A distribution of the 172 *average* scores of the individual schools showed a standard deviation of 15.81 score units (column 4). The variability of raw scores of individual pupils was therefore (column 5) only 1.66 times as large as the variability in school averages. In other words, the differences between schools were of almost the same magnitude as the difference between individuals in all schools.

It will be noted (column 5) that in only four of the tests was the variability in raw scores more than twice as large as the variability in school averages, and that the average ratio of these variabilities is slightly less than 2 for the ten tests. It should also be noted that all of the separate ratios are surprisingly close to this average value. A glance at columns 1 and 3 should be sufficient to indicate the statis-

tical significance and reliability of these findings. To the extent that the tests used are typical of current standardized tests, it may be stated as a fair generalization that the variability in *school performance* on standardized high school tests is related to the variability of achievement of individual pupils in the approximate ratio of 1:2. For the total 53,669 pupil-testings, the average number of pupils per test per school was 29. In a series of *random* samples of this size the ratio between variability in scores and in averages would be $1:\sqrt{29} = 1:5.4$. Clearly, therefore, the variability in school averages is far greater than can be accounted for by chance factors alone. This variability, furthermore, is not primarily due to the influence of schools of small enrollment, as might well be supposed. A study of average performances of schools classified into three enrollment groups showed that the large schools (over five hundred in enrollment) showed only slightly less variability in average scores than did the small schools (less than one hundred twenty-five in enrollment). In other words, the variability in school averages, contrary to expectation, appears to be almost independent of the number of pupils tested in each school.

Significance of School Differences.—What is the significance of these facts in the problem of establishing adequately reliable norms for standardized tests? The first point to be made is that the *number of schools* sampled is of far greater importance than the number of individuals tested; so much more important, in fact, that the latter need not be at all considered, since a large sample will automatically be provided if the number of schools sampled is sufficiently large and well selected. With such relatively large differences existing from school to school, a norm based upon scores obtained from a small number of schools, *regardless of the number of scores* obtained, is likely to deviate so far from the mean of the population that little if any reliance may be placed upon it. It may be demonstrated, on the basis of the data presented, that a norm based upon a sample of from five or six schools (true of some current norms) is statistically no more reliable than one based upon the scores of 20 to 25 pupils, if these pupils are selected truly at random from the entire population.

Because of the fact, therefore, that a population of individuals grouped into schools showing such great variability is not a population statistically, *i.e.*, is not fundamentally homogeneous, *it becomes essential that the school, rather than the pupil, be considered as the basic unit.* For practical, as well as theoretical considerations, furthermore, it is absolutely essential that the school be considered as the basic unit.

It is obviously impracticable to attempt to secure a *strictly random* sample from an entire population of school children. Schools must be tested as schools if any testing is to be done at all; one cannot go into all schools to test only one or two pupils selected at random in each. For the purpose of establishing reliable norms, therefore, the population must be considered as one of the schools, not one of individuals.

Considering then the school as the basic unit, let us attempt to determine empirically the number of schools that must be tested in order to establish an adequately reliable norm. According to sampling error theory, the probable error of a mean obtained from a random sample is in direct proportion to the variability of the population from which the sample is drawn, and in inverse proportion to the number of cases in the sample. The greater the variability in the population, therefore, the larger is the sample required to secure a given degree of reliability. Obviously, with the variability in schools as large as it is, the number of schools required is likely to prove greater than has ordinarily been supposed,—at least greater than has usually been secured in practice.

A Tentative Standard of Norm Reliability.—There does not exist at the present time any generally accepted criterion or standard of reliability for high school test norms—any agreement upon what should constitute the maximum amount of probable sampling error permissible in such a norm. In order to make the remaining discussion specific, therefore, the writer is forced to suggest, arbitrarily, an objective standard. No claim need be made for this standard, since it is proposed primarily only for the purpose of illustration to facilitate this discussion. Let us set up, therefore, as a tentative standard, the requirement that the standard error of a norm should not exceed one-eighth of the standard deviation of the population (of school averages) that it represents. In terms of the 2:1 ratio previously established, this is in general equivalent to requiring that the standard error of the norm may not exceed one-sixteenth of the standard deviation of raw scores (of individuals) in the entire population. Otherwise expressed, it is the degree of reliability that would result from a sampling of two hundred fifty-six pupils, if these pupils were selected strictly at random from the entire population. Speaking again in terms of school average, it only *insures* (within limits of practical certainty) that the norm is not in error (due to sampling) by more than three-eighths of the standard deviation of obtained school average.

Expressed in such terms, it certainly does not appear that the tentative criterion is unreasonably high—rather the contrary.

How many schools must be tested in standardizing a test in order that the norm may satisfy this criterion of reliability? If the schools taken are truly selected at random, the ratio of the standard error of the norm to the standard deviation of obtained school averages is equal to the square root of the number of schools sampled. To meet the criterion of a ratio of 1:8, therefore, it is necessary that *sixty-four schools* be tested. It cannot be too strongly emphasized, also, that these sixty-four schools must represent a *random* selection from all schools in the population.

How many norms now available can satisfy this reasonable criterion? If the reader is at all informed concerning current tests, he can answer this question to his own satisfaction, without any further presentation of facts here. When one considers not only the tendency to be satisfied with large samples from a small number of schools, but the fact that the test constructor tends to secure most of his samples from large schools (since a large sample can thus most conveniently be obtained), from progressive school systems (since such systems are most ready to cooperate), and from schools within a limited geographical area (again because of convenience), the case for available norms is even more seriously weakened.

It might be well to emphasize the fact that all of the empirical ratios and estimates derived in the preceding paragraphs are based upon facts for the State of Iowa, and may strictly be applied only to that state. Rigid adherence to a state course of study and to uniform standards of teacher certification and training would tend to lessen these differences. Lack of uniformity in content and method, however, is still typical of high school instruction in most of the states. The same type of evidence secured elsewhere would very likely lead to similar conclusions and to quantitative estimates of the same order of magnitude.

Need for Norms of Variability in School Averages.—When a standardized test is administered for the purpose of evaluating instruction in a given school, the question to be answered is not "How does the average performance in this school compare with the performance of individual pupils?" but, rather "How does the average performance of this school compare with the *average performances* of *other schools?*" An adequate answer to the latter question requires that norms be expressed in terms of school averages as well as in terms of raw scores.

With almost no exception, however, it is only the latter type of norm that has been made available. The additional norms required could perhaps be best provided by presenting the complete frequency distribution of *school averages* for the schools on which the other norms are established. For the sake of convenience of interpretation, percentiles or measures of variability should be computed for this distribution. On the basis of such norms it would be possible to make a statement such as the following: "School A showed an average performance on Test I that is exceeded only by fifteen out of every one hundred schools in the state." No such statement can be made on the basis of norms in terms of raw scores, since there is no strictly constant ratio between variability in raw scores and variability in school averages.

Need for Norms for Several Specific Populations.—The preceding discussions and estimates have all been based on and computed for a population which might be defined as "*all* the high schools in the State of Iowa." It is doubtful, however, that norms established for a population thus defined would alone be sufficient, or that they are even the types of norms which would prove most valuable and meaningful to a school administrator interested in determining the relative effectiveness of instruction in his own school. All schools within this general population are not comparable. The large schools in the Iowa Academic Contest, for example, secured higher mean scores, on the average, than did the small rural schools. School B, a large city high school, would perhaps be primarily interested in a comparison of its mean performance with those of other large city high schools, not with the general norm for the entire state.

This suggests that in addition to the general type of norm, there should be provided separate norms for each group of schools classified on the basis of enrollment.

Systematic differences between schools due to or related to the factor of enrollment, however, are only one of a number of possible bases for the definition of specific sub-populations for which separate norms might be established. Differences due to the use of different texts or courses of study, to different methods of teaching, to differences in length of school year, to different types of school organization, and to geographical differences, are among some that might be mentioned.

It is not seriously proposed that specific norms be established for all or any tests on the basis of all the factors mentioned. While such

norms would surely be of value in the interpretation of school results, it is doubtful that they would all be worth the trouble involved in securing them. The important point to be made is that such systematic differences do exist, and that, in proportion to their relative magnitude, they operate to detract from the meaning and value of the general type of norm now available.

Can and Should Adequate Norms Be Established?—The answer to the first part of the question is—yes, but not by single individuals or by single institutions. The extensive and representative sample required to establish an adequate *national* norm, for example, can perhaps only be secured through the functioning of a centralized, national, noncommercialistic agency, designed to secure and coordinate the cooperation of a large number of institutions now interested and engaged in test construction. The services of such an agency should and could be made available only for the standardization of the most carefully constructed and validated test materials, upon the approval of a committee made up of the country's leading experts in the field on educational testing.

Are adequate norms for high school tests "worth the candle"? That is a question that the writer prefers to leave entirely to the reader, since it is largely a matter of opinion. One point, however, may be safely made. If norms are worth establishing at all, they are worth establishing adequately. The norms now available for many tests are perhaps no better than no norms at all. They may even be as often *seriously misleading* to the user as they are highly accurate. Either we must insist that norms be established and expressed in such a manner that the user may be reasonably certain of their meaning and reliability, or we must admit that their value does not justify their cost, and therefore do without them rather than be misled by defective substitutes.

Let it be said in conclusion that this article has been deliberately designed to challenge the complacency of those who have become too uncritical of the present status of educational tests. For that reason not as much credit has been given to the very few well-standardized high school tests now available that such tests deserve. It must be noted also that the charges made apply primarily to tests in the secondary school field. It is not likely that school differences are of comparable magnitude and significance for the fundamental subjects in the elementary school, although it is quite possible that even in the

elementary field they are of greater significance than practice has recognized. Until uniformity in content and procedure of high school construction becomes the rule, rather than the exception, therefore, the task of establishing norms for high school tests cannot and should not be left to the limited resources of the M.A. or Ph.D. candidate, or of any single individual or institution.

HOW SUCCESSFUL COLLEGE STUDENTS STUDY

JAMES D. WEINLAND

New York University

Efficiency in work and study is very much to be desired. In consequence the literature has become filled with the account of methods to that end which are so numerous that they are hard to evaluate. It is easy to suggest dozens of things which a person should do in order to succeed and the poor student is very apt to be discouraged and dismayed by the prolixity of the advice. All of the suggestions have probably proved valuable under some circumstances to some people. But which ones have the widest application and are the most important?

This investigation was begun to find out just how successful college students do their work and to discover the methods and particulars upon which there is the most agreement. A questionnaire of seventy items was prepared, the questions, for the most part, being selected from the various discussions on how to study; Book, Kitson, McClure and others. These questionnaires were presented to honor students with the request that they answer the questions. All were men, upper classmen in the School of Commerce of New York University. In every case an interview was held with the student answering the questions to make as sure as possible that the attempt to answer the questions would be honestly and sincerely made. In a number of cases where the cooperation did not appear to be good the questionnaires were rejected. Twenty men finally answered the questions and their answers, together with some impressions from the interviews form the basis of this study. Five students who failed also filled in the questionnaires and in a few cases where their answers most uniformly disagreed with the answers of the successful students, the contrasts are given.

The questionnaire consisted of seventy items of which the first thirteen are reproduced. The queries which followed covered such points and as how, when, and where study was conducted, methods of note-taking and the use of notes, attitudes in class, etc. Half of these men were specializing in banking and finance and half of them in economics. We have made nothing of that fact here though the original plan calls for the collection of data from different curricula to determine whether different methods of work (note taking, etc.) apply better to different courses of study. It is quite probable, however, that the

fact that all but one of the men had selected their vocation was due to the fact that we took our questionnaires only to men who were specializing in one of the two subjects named above. Many successful students in the general courses, we believe, have little idea of the vocation which they will follow. In consequence, even though we believe the possession of a definite purpose to be of great value, we have not stressed the selection of a vocation in our conclusions.

In some cases such as question 6 (What kind of exercise do you take?) etc., where the answers were quite varied and without particular point, so far as we have seen, we have not included the answers here. The first thirteen items on questionnaire follow:

Specialization

Questionnaire on Methods of Study

Name Age Sex

Previous education

Physical

1. Meals Heavy light none
   Morning.....
   noon.....
   evening.....
   late at night
2. Stimulants; coffee.....tea.....others.....
3. Sleep Hours When regular?
4. Cigarettes how many when
5. exercise hours when regular
6. exercise; what kinds.....
7. Do you have any rules for keeping yourself in good physical condition?
.....

Purpose

8. Have you selected your vocation?.....
9. Are your courses in line with your vocation?.....
10. Are you really interested in your subjects?.....
11. Can you sustain interest through a course?.....or is it usually high in spots?.....
12. Do you have hobbies?.....
13. Are any of your courses in line with your hobbies?.....

Results

All the students did not answer all the questions, as in some cases (hours of sleep, etc.) some were not sure, and in other cases (Do you follow a mental schedule, etc.?) a few of them did not understand the meaning of the question. Furthermore we pressed more strictly

for sincerity of response than completeness. The answers as given, however, made possible some weighting of the importance of the different statements and below they are given in order of merit, with the most emphasized and agreed upon factors heading the list and the least agreed upon factors at the end.

We included in the following list methods advised in the literature but not used by these particular students. In doing this we do not wish to infer that such methods might not be useful, but to register this particular vote as to their importance. It has been the writer's experience that in listening to advice on how to study he has heard certain things very much emphasized because the method has helped the speaker, or has been proven in some laboratory experiment, etc. He has himself emphasized certain things such as the timing of reviews and the marking of books that these twenty successful students did not in the least agree upon. Of course, because good students did not use these methods is no indication that they would not have been better students if they had, but it should be a safe plan in advising poor students to emphasize first of all the methods actually in use by the greatest number of successful students. Once these fundamentals are mastered, refinements of method might be added. In brief, since advice must be pointed to be helpful, it is better to emphasize methods used by a number of good students than some method found helpful in one's own individual experience or accepted for theoretical reasons. Laboratory experiments in learning may prove definite things but when the conclusions are transferred to the learning done by college students in the midst of complicated living they may lose, or gain, a great deal in importance. It is safest to emphasize first the methods used by successful students, at work.

ORDER OF MERIT RATING OF IMPORTANT FACTORS IN STUDY AS USED BY TWENTY HONOR STUDENTS

1. Can work well alone.
2. Have selected their vocation.
3. Courses are in line with their vocation.
4. Do not study in groups where tobacco smoke is thick.
5. Rest when tired rather than work right ahead.
6. Avoid all distractions while studying.
7. Work by assignment.
8. They study alone.
9. They are fast readers (read phrases at a glance).
10. Stop occasionally in their reading to think over the subject and make sure they grasp it.

11. Recall and think over at night, or other time, what they have been reading.
12. They do not leave space in their notes to fill in new ideas that come to them.
13. They keep full notes.
14. They follow up suggestions made in class.
15. They tend to question authority.
16. They have hobbies (not necessarily in line with their studies).
17. They note the examples as well as the principles of their instructors.
18. They do not review their notes before going on to a new assignment each day.
19. They dance, play cards, belong to social groups.
20. They eat a light breakfast, a light lunch, a heavy dinner in the evening.
21. They take from seven to eight hours sleep.
22. They use coffee or tea stimulants.
23. They talk with other students in class.
24. They are really interested in their subjects.
25. They make systematic review before examinations.
26. They have confidence in their memories.
27. They study their instructors methods.
28. They look up things not clear in class.
29. They concentrate easily.
30. They ask other people about things not clear in class.
31. They study idiosyncrasies of instructor and take advantage of them.
32. They follow a mental schedule in their work.
33. They do not study in class.
34. They make mental inventory before reading a book.
35. They read intensly (once).
36. They get much from the class period.
37. They do not re-write their notes.
38. They cram for examinations.
39. They take part in athletics.
40. They can sustain interest uniformly throughout a course.
41. They do much miscellaneous reading.
42. They ask questions in class.

Before drawing any conclusions it will be well to note some of the strongest comparisons from the questionnaires of the "flunk" students. They agree somewhat in the following (although these items are not put in order of merit as there is not enough reliability in the five cases to make this worthwhile):

They do not strictly avoid distractions while studying.
Their notes are not so full.
They cannot work on a subject which does not interest them.
Sometimes they study with a group where the tobacco smoke is thick.
They ask no questions in class.
They work right on when tired.
They cannot sustain interest very well throughout a course.
They eat late at night and have no rules for keeping physically fit.

Conclusions

An examination of the order of merit list of methods used or neglected by the successful students and the short list of comments from the unsuccessful students will enable the reader to draw his own conclusions.

To the writer the following points seem of major importance:

1. All of the good students were able to work alone and most of them did study alone. This may indicate either self-reliance or purpose. At any rate the initiative required to attack one's own job and do it seems of prime importance. A multitude of suggestions may come from group cooperation but the class situation gives scope to this factor, and afterward individual self-reliant effort seems to mark the good student.

2. The avoidance of distractions, particularly noise stimuli, is important.

3. Successful students make it a practice not to work when tired. Short rest periods more than pay for themselves. Alternating tasks proves refreshing. A certain ability to manage time is as important as good intentions.

4. Most of the good students worked by assignment; they were able, to some extent, both to sustain interest and to work on when interest was lacking. We may conclude from this that they have a certain valuable power of control. Some students, for instance, have great difficulty in confining themselves to work on assignments and tend to read miscellaneously, examining whatever they are interested in, rather than finishing the work at hand. This tendency, if allied with enough ability, may perhaps produce genius, but in the average college student the ability to limit and direct effort to the problems assigned by the instructor seems to be more valuable.

5. Good students keep fairly complete notes. The way they use them seems to vary a good deal, although a systematic review before examinations is quite common. Cramming is considered valuable when necessary.

6. Successful students have enough interest or purpose in their work to recall and think over what they have heard and read. Organization of the material is furthered and the memory strengthened by this process.

7. Usually the good students eat lightly in the fore part of the day and heavier in the evening. Definite food habits are helpful in organiz-

ing time and directing effort. Interviews have disclosed that many students, especially freshmen, are at a low level of efficiency because, having left home, their living habits become disorganized and they have not been able to establish new ones. The poor students responded one hundred per cent that they had no general rules for keeping fit.

8. The ability to read rapidly and the general confidence displayed would indicate that success in college is built upon previous successes and skills already developed.

9. Most of the good students were social. They talk with each other both in class and out. They belong to social groups and usually dance or play cards. Perhaps the fact that they have hobbies should be listed here. They are not often of the shut-in type.

10. There is considerable divergence in particular ways of doing things to be found among them. Most of them do not leave space in their notes for the addition of new ideas that come to them, they do not make a practice of reviewing each day before going on to the new assignment, some mark their books and some do not, etc. We may conclude, I think, that though these methods are good they are not fundamental.

11. It is particularly worthwhile for instructors to know that good students observe and note their examples and illustrations with as much attention as they give to the principles that are being discussed. It is the illustration which the student can most easily remember and he will attach the principle to that. Great care should be expended in making sure that the illustrations are pertinent and really are examples of the principle which is being taught.

12. In interviews with students regarding methods of study the exclamation has frequently been made; "Why wasn't this told us when we were freshmen?" Methods must become habits before they are very useful and it takes time to develop habits. If beginning students were given more training in how to study more efficient graduates would be produced.

THE POETRY OF PROMISE[1]

ROSE SUSAN MALMUD

London, England

Aeschylus: What I want is to bring him to the balance;
The one sure test of what our art is worth!
Dionysus: So that's my business next? Come forward, please;
I'll weigh out poetry like so much cheese![2]

At the risk of having these mocking lines from Aristophanes' "Frogs" hurled at me, I shall attempt to present the subject of measurement in relation to what has always been thought and may continue to be an incommensurable.

Accounts of the history of measurement of individual differences tell of its inception at the end of the nineteenth century, but there is undoubtedly evidence that tests of individual differences were in use as far back as the time of Gideon,[3] if not before. His method of rejecting warriors who knelt to drink might be accepted as effective even today, for surely those who kneel to drink cannot be on the *qui vive* to the same extent as those who do not kneel. The end of the nineteenth century marks rather the beginning of scepticism in measurement and in these terms only would the writer wish to be regarded as part of the movement.

If in criticism, we have been persuaded to accept man as the measure of things literary, we have legitimately asked "which man" and then just as legitimately taken his measure with more or less ready adjectives of praise, wonder, or condemnation. Whether we have succeeded by this means in reducing the number of unknowns, always a major problem in criticism, is questionable. At any rate the intention to do so was there. Coleridge, in Chap. XV of his "Biographia Literaria,"[4] on the basis of an analysis of Shakespeare's earliest known works made generalizations on what constitutes promise in the writings of young poets. According to him, there are four main criteria:

---

[1] The material in this article was gathered during the tenure of a fellowship from the Social Science Research Council. The writer takes this opportunity of expressing appreciation to those who in the course of the study rendered valuable clerical assistance, particularly to Miss Jean Brown and to Miss Helen Malmud.

[2] Translated by Gilbert Murray.

[3] Judges: VII, 3–7: "The people are yet too many; bring them down unto the water, and I will try them for thee there."

[4] Coleridge, S. T.: "Biographia Literaria," prepared for the public by Henry Nelson Coleridge. New York: William L. Allison, 1882, Chap. XV, pp. 453–459.

1. The delight in richness and sweetness of sound, even to a faulty excess, if it be evidently original, and not the result of an easily imitable mechanism, I regard as a highly favorable promise in the compositions of a young man.

2. A second promise of genius is the choice of subjects very remote from the private interests and circumstances of the writer himself.

3. Images "become proofs of original genius only so far as they are modified by a predominant passion; or by associated thoughts or images awakened by that passion; or when they have the effect of reducing multitude to unity, or succession to an instant; or, lastly when a human and intellectual life is transferred to them from the poet's own spirit."

4. A fourth promise of genius is depth and energy of thought. "No man was ever yet a great poet, without being at the same time a profound philosopher."

It may be true that fine literary truths can be captured in the meshes of a fine mind only; we no doubt need a Coleridge to evaluate a Shakespeare. Nevertheless, that necessity should not blind us to the possible fallacies involved. Coleridge studied Shakespeare's early writing after he had been nourished by the fruit of Shakespeare's mature genius. The aura of greatness which is projected to a recognized poet's early work sometimes transfigures criticism strangely. Even if our emotion for the mature poetry did not disarm hostile criticism of the youthful writing, the inescapable fact of the poet's achievement renders slightly ridiculous any unfavorable criticism directed toward his early efforts. Then, too, may we accept unquestionably generalizations on the work of one supreme poet at a particular stage of his growth as valid for all poets at any stage? Involved in the problem are the assumptions: (1) That poetic gifts manifest themselves and are recognizable in very early youth, as in the acorn we may identify the oak; (2) that in order better to understand poetic gifts, it is important to study in innumerable instances the directions in which they vary. We use the term "poetic gifts" advisedly for many variables are involved: The perception of rhythmic patterns or kinæsthetic sensitivity; the perception of the identity or difference of vowal and consonantal patterns; the perception of social situations—perceptions formed by previous affective experiences[1]—all the foregoing being constituents of literary perception itself, or the sensitivity to the highly selected verbal meanings which re-create experience.

Three avenues of study are open: (1) The poetry of contemporary children who have had verse published in school magazines, books, or newspapers; (2) the early poetry of the poets and critics who have

[1]See Poetry and the Emotions, an article by the writer in *The Journal of Abnormal and Social Psychology*, Vol. XXII, No. 4, January–March, 1928.

established themselves in literary history; (3) tests of literary and social perception to throw light on the nature of literary gifts and their relation to literary creation.

In connection with the first, a preliminary survey was made of verse written by children in elementary and secondary school magazines[1] as well as in newspapers covering a period of approximately two years. The total number of poems that the writer found it thus possible to secure were 1917, representing twenty-six states in the United States and two hundred nine cities, including twelve cities in England and Canada. Of this total, one hundred eighty-five poems were selected, for objective[2] analysis. These were the shorter poems ranging from four to twenty lines. It is probable that by this means a wider variety of meaning and sound patterns was made available than would have been possible from a random choice of a hundred or two hundred from the total number. Furthermore since far fewer elementary school magazines were found, the selection of shorter poems gave better representation to the writings of younger children. In the face of the popular conception that statistics means standardization, the table of objective facts is presented with a view to emphasizing our criterion of significance, which is departure from the average.

Total number of poems surveyed	1917
Sources: elementary and high school magazines, newspapers:	
States represented	26
Cities represented (Including twelve from England and Canada)	209
Number of short poems selected for analysis (four to twenty lines)	185
Semantic aspects of the analysis:	
	WORDS
Range of vocabulary (counting only nouns, verbs, adjectives and adverbs)	6 to 87
Average vocabulary (decimals have been dropped)	31 σ 14
Average number of nouns	12 σ 6
Average number of verbs	9 σ 4
Average number of adjectives	7 σ 4
Average number of adverbs	3 σ 2
Types:	
	PER CENT
Subjective poems (apparently personal experience)	44
Objective poems	56
On aspects of nature	51
(As opposed to twenty-six per cent of the total number of poems on the same theme)	

[1] The writer wishes to express gratitude to Joseph M. Murphy, editor of *The School Press Review*, who made possible the examination of the greater part of the magazines.

[2] By objective, we mean verifiable.

	Per Cent
Treatment:	
Descriptive	78
Interpretative	22
Use of figurative expression:	
One or more metaphors or similes by	27
Two or more metaphors or similes by	13
Three or more metaphors or similes by	3
Personification by those who used no other type of figurative expression	16
Metonymy by those who used no other type	1
Phonetic aspects of the analysis:	
Use of rhythm:	
Iambic meter	70
Perfectly regular iambic meter	8
Trochaic meter	13
Perfectly regular trochaic meter	2
Anapestic meter	13
Perfectly regular anapestic meter	1
Dactyllic meter	4
Use of line patterns:	
Tetrameter	32
Tetrameter and trimeter	32
Trimeter	9
Trimeter and pentameter	4
Pentameter	4
Pentameter, tetrameter and trimeter	4
Dimeter, trimeter and tetrameter	4
Dimeter and trimeter	4
Dimeter	3
Dimeter and tetrameter	3
Various other combinations of meters	1
Use of rhyme:	
Number of unrhymed poems	4
Consecutive rhymes (aabb, etc.)	40
Alternate rhymes (abab, etc.)	41
Crudely mixed rhymes	6
Skillfully mixed rhymes	8
Number of rhymes:	
One or more	96
Two or more	84
Three or more	52
Four or more	36
Five or more	10
Six or more	6
Seven or more	3
Eight or more	1

	Per Cent
Use of alliteration:	
One or more times	78
Two or more times	50
Three or more times	31
Four or more times	17
Five or more times	13
Six or more times	9
Seven or more times	5
Eight or more times	3
Nine or more times	2
Absence of alliteration	22
Use of assonance (vowal agreement):	
One or more times	85
Two or more times	81
Four or more times	53
Six or more times	38
Eight or more times	22
Ten or more times	14
Twelve or more times	8
Fourteen or more times	5
Sixteen or more times	4
Eighteen or more times	3
Twenty or more times	2
Absence of assonance	15
Use of consonance (consonantal agreement):	
One or more times	73
Two or more times	67
Four or more times	43
Six or more times	25
Eight or more times	14
Twelve or more times	7
Fourteen or more times	4
Sixteen or more times	2
Absence of consonance	27

With this preliminary discipline in criticism, analysis of a special group of poems may be profitably pursued. In order that the semantic and phonetic variables may be the more easily observable, the subject "Snow" has been kept constant. The ages of the children range approximately from seven to eighteen; the distribution according to school is as follows:

Elementary School	High School	Magazine
1. E. L. (Grade II)		
2. J. P. (Grade VII)		The Cortelyan
3.	P. M.	
6.	T. D.	Sherrill High School Spirit
7. B. T. (Grade V)		The Quill
9. J. K. (Grade VI)		Ye Horace Manuscript
10.	S. B.	De Witt Clinton
13.	A. E. H.	Nardin Quarterly

The series ends with Francis Thompson's poem, an inclusion suggested by a comment made by T. Sturge Moore in connection with a previous study of a miscellaneous group of children's poems. The poems have been arranged in tentative order of complexity, consideration being duly given to the significant facts in accordance with the table.

While it may be true that the perfect poem is comprehended as a unity, music and meaning forming an integrated whole, in the imperfect one which must be judged not for what it tells but for what it foretells, sound and sense are independent variables and may be analyzed as such. Unity is a growth and may be observed in process of evolution in the steps represented by the poems. The implication is not that Thompson's poem is the perfect poem on the subject of snow for with newer conceptions of time and space in relation to each other and to man, a poet's ascent or descent in an aeroplane during a snowfall, for example, may yield the perfect facet of human experience of snow.

Selection tends in the direction of the more concrete and the greater contrast effect; from adjectives implying vague, general characteristics in a state of being, the choice is transferred to the pungently concrete adjective and the adverb emphasizing action and finally process.

Simultaneously with the sharpening of outline due to the narrowing of choice of descriptive word, there is a broadening of the canvas; from the relation of the child to snow and to play, the transition is to greater attention to the characteristics of snow and its message and finally to the identification of its characteristics with the process of formation by the Creator. The more complex stages of poem evolution show figures of speech reinforcing one another. Furthermore, the observation and inference necessary in the creation of effective figures of speech seems to operate as well in the selection of suitable sound patterns. Thus, in the most complex poem, the items of semantic

and phonetic analysis truly become aspects of one solid structure and not as in the early stages of poem evolution, lines that intersect at only a given point.

1

1. Snow is white
2. It comes down one, two, three
3. It says that all the children
4. Will love to play with me.

Line number	Semantic aspects, figures of speech	Phontic aspects		
		Rhyme	Rhythm	
			Line	Foot
1			dimeter	trochee
2		assonance	trimeter	iambic
3	Personification		trimeter	iambic
4		rhyme	trimeter	iambic

Type, descriptive.
Number of words, 21.

2

Snow

1. Snow, snow, beautiful snow,
2. Drifting along the sky,
3. 'Tis nature's soft and lovely snow,
4. That blows from house-tops high.

Line number	Semantic aspects, figures of speech	Phontic aspects		
		Rhyme	Rhythm	
			Line	Foot
1		repetition	trimeter	spondee dactyl
2		alliteration	trimeter	dactyl iambic
3		alliteration	tetrameter	iambic
4		rhyme	trimeter	iambic

Type, descriptive.
Number of words, 19.

3

The Snow

1. The snow is falling thick and fast
2. As it has for many long years past.
3. The wind is hurrying it soft to the ground
4. Where it alights without making a sound.
5. The children, from my window, I can see
6. Are enjoying themselves, in utmost glee.
7. The beauty of it they don't seem to know
8. But are making the best of the beautiful snow.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1			tetrameter	iambic
2		rhyme	tetrameter	anapest iambic
3			pentameter	iambic anapest
4	Personification	rhyme	pentameter	iambic
5		assonance consonance	pentameter	iambic
6		rhyme	tetrameter	iambic anapest
7		consonance assonance	tetrameter	iambic anapest
8		rhyme consonance	tetrameter	iambic anapest

Type, descriptive.
Number of words, 63.

6[1]

A Snowflake

1. We watched you as you slowly fell
2. Upon the cold, hard ground;
3. Changing hill and vale and dell
4. To white without a sound.

5. You fell to earth in dead of night
6. Or in the midst of day,
7. Children ran screaming with delight
8. Along your white pathway.

[1] The gaps between numbers indicate omission of poems because of lack of space.

9. What is Christmas without snow?
10. No place for skii or sled,
11. So you are not forgotten, though
12. In summer you are dead.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1	Personification		tetrameter	iambic
2		assonance consonance	trimeter	iambic
3		rhyme assonance consonance	tetrameter	trochee
4		rhyme consonance assonance	trimeter	iambic
5		assonance	tetrameter	iambic
6		assonance	trimeter	iambic
7		assonance rhyme	tetrameter	iambic trochee
8		rhyme rhyme	trimeter	iambic
9		assonance	tetrameter	trochee
10		alliteration consonance	trimeter	iambic
11		rhyme rhyme	tetrameter	iambic
12		rhyme	trimeter	iambic

Type, descriptive,
Number of words, 68.

7

The Snowflake

1. I am a little snowflake,
2. So fluffy and so white.
3. I flutter, flutter, through the air,
4. And softly I alight.
5. The great moon shines so brightly,
6. It makes the whole world glow.
7. It seems to me like fairy land,
8. When covered with the snow.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1	Personification		trimeter	iambic
2			trimeter	iambic
3	Onomatopoeia	assonance alliteration repetition	tetrameter	iambic
4		consonance rime	trimeter	iambic
5		assonance consonance	trimeter	iambic
6		assonance assonance	trimeter	iambic
7	Simile	assonance	tetrameter	iambic
8		rhyme	trimeter	iambic

Type, descriptive.
Number of words, 44.

9

The Snow Drop

1. Small and white
2. Yet brave,
3. A lamb with lion's heart,
4. A hero clothed in white,
5. A messenger to tell us
6. Spring is on her way.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1			dimeter	trochee
2			monometer	iambic
3	Metaphor	alliteration	trimeter	iambic
4	Metaphor		trimeter	iambic
5	Metaphor	assonance	trimeter	iambic
6	Personification		trimeter	iambic

Type, interpretative.
Number of words, 25.

10

Melting Snow

1. How beautiful it is, melting snow.
2. Glistening, ever lessening in the sun.
3. As I watch your consistency go
4. Into mystery.

5. How pathetic it is, melting snow.
6. Whose battlements of yesterday
7. Are now the footstools of the low
8. And despised.

9. How weak it is, melting snow.
10. Just as a million people live,
11. Who were strong; but then they go
12. Into earth.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1			tetrameter	iambic anapest
2		assonance assonance consonance	tetrameter	dactyl trochee anapest
3		rhyme	trimeter	anapest
4		assonance consonance	dimeter	iambic
5		repetition	trimeter	anapestic iambic
6	Metaphor	consonance	tetrameter	iambic
7	Metaphor	rhyme	tetrameter	iambic
8			monometer	anapestic
9		repetition	trimeter	iambic
10	Simile	assonance	tetrameter	iambic
11		rhyme	trimeter	anapestic iambic
12			monometer	anapestic

Type, Interpretative.
Number of words, 60.

13

A Snowflake

1. A bit of white from a heavenly home
2. Comes drifting down to earth:
3. As dainty fresh, enticing sweet
4. As a baby's soul at birth.

5. So innocent—this wisp of down
6. That floats on frigid air,
7. One almost thinks it might have been
8. An angel's morning prayer.

9. Then slowly with a fairy grace
10. Alighting on the ground,
11. It mingles with its fellow flakes
12. And never more is found;

13. Until all bruised and ugly too,
14. And to the dark earth pressed,
15. It begs a kindly ray of sun
16. To send it back to rest.

17. Our souls were like these flakes of snow,
18. Sent forth by heaven too,
19. And, like the snowflakes, pure and white
20. Until the earth they knew.

21. And now by the wretched world befouled,
22. Their beauty lost, they strive
23. To find abode with Him above
24. Whose breath made them alive.

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1		alliteration consonance	tetrameter	iambic anapest
2		alliteration	trimeter	iambic
3		consonance	tetrameter	iambic
4	Simile	alliteration rhyme	trimeter	iambic anapest
5	Metaphor	assonance consonance	tetrameter	iambic iambic
6		alliteration	trimeter	iambic
7			tetrameter	iambic anapest
8	Metaphor	rhyme	trimeter	iambic
9		assonance	tetrameter	iambic anapest
10	Personification		trimeter	iambic
11		assonance alliteration	tetrameter	iambic
12		assonance rhyme alliteration	trimeter	iambic
13		assonance	tetrameter	iambic
14		repetition	trimeter	iambic
15		assonance	tetrameter	iambic
16		alliteration rhyme	trimeter	iambic
17	Simile	alliteration assonance	tetrameter	iambic
18		alliteration	trimeter	iambic
19		repetition assonance	tetrameter	iambic
20		rhyme	trimeter	iambic
21		assonance consonance	tetrameter	iambic anapest
22		consonance	trimeter	iambic
23		assonance consonance alliteration	tetrameter	iambic
24		alliteration rhyme assonance consonance	trimeter	iambic

Type, interpretative.
Number of words, 141.

15

To a Snowflake

1. What heart could have thought you?
2. Past our devisal
3. (O filigree petal!)
4. Fashioned so purely,
5. Fragilely, surely,
6. From what Paradisal
7. Imagineless metal,
8. Too costly for cost?
9. Who hammered you, wrought you,
10. From argentine vapour?
11. 'God was my shaper
12. Passing surmisal,
13. He hammered, He wrought me,
14. From curled silver vapour,
15. To lust of His mind:–
16. Thou could'st not have thought me!
17. So purely, so palely,
18. Tinily, surely,
19. Mightily, frailly,
20. Insculped and embossed,
21. With His hammer of wind,
22. And His graver of frost.'

Line number	Semantic aspects, figures of speech	Phonetic aspects		
		Rhyme	Rhythm	
			Line	Foot
1		consonance	dimeter	iambic anapest
2		assonance[1]	dimeter	iambic[2] anapest
		consonance		
3	Metaphor	alliteration	dimeter	anapest
		consonance		
4		alliteration	dimeter	iambic anapest
		alliteration		
5		alliteration	dimeter	iambic anapest
		rhyme		
		assonance		
6			dimeter	anapest
7	Metaphor	rhyme	dimeter	anapest
		alliteration		
8		repetition	dimeter	anapest
		consonance		
9	Metaphor	rhyme	dimeter	iambic anapest
		consonance		
10			dimeter	anapest
11	Personification	rhyme	dimeter	iambic anapest
12		rhyme	dimeter	iambic anapest
13		repetition	dimeter	anapest
14		repetition	dimeter	anapest
15			dimeter	anapest
16		repetition	dimeter	iambic anapest
17		repetition	dimeter	anapest
		alliteration		
18		repetition	dimeter	iambic anapest
		rhyme		
19		rhyme	dimeter	iambic anapest
		assonance		iambic anapest
20	Metaphor	rhyme	dimeter	anapest
21	Metaphor	rhyme	dimeter	anapest
22	Metaphor	rhyme	dimeter	anapest

Type, interpretative.
Number of words, 81.

[1] Pronounciations are based on "An English Pronouncing Dictionary" by Daniel Jones. New York: E. P. Dutton and Co., 1924.

[2] The last unaccented syllable in the first line is carried over to form an iambic in the second line; similarly the last unaccented syllable in the second line is carried over to form an anapest in the third line.

Summary and Conclusions

In the analysis of the semantic aspects of these short poems, only twenty-two per cent were found to be interpretative in character as opposed to seventy-eight per cent purely descriptive. Inasmuch as Coleridge offered no clue as to how depth of thought may be observed or compared as between poems, this tentative classification may be accepted as an aid, for surely interpretative poems penetrate more layers of thought than purely descriptive ones. We say purely descriptive ones, for if only one line departs from the descriptive type, it can no longer be classified as such. Keats, for example, is thought of by many as being a purely descriptive poet of nature in the face of such lines as:

> Thou wast not born for death, immortal Bird!
> No hungry generations tread thee down.

However Coleridge's criterion of objectivity as choice of subjects remote from the writer's private interests is not supported by our facts. Writing objectively is somewhat more common (fifty-six per cent), and apparently somewhat less difficult than writing subjectively. However an examination of the longer poems may change the relations. It is interesting to note in this connection that nature poems formed fifty-one per cent of the shorter poems but only twenty-six per cent of the total number of poems. This criterion, in any case, is open to criticism, for the mere writing in the first person is no proof that the individual is not identifying himself with another, nor does his writing in the third person preclude the possibility of his projecting his own experiences. Aristotle's conviction[1] that the use of figurative language is the most important mark of the poet seems to be borne out by our significant facts. In only twenty-seven per cent of the poems did we find the use of one or more similes or metaphors; in only thirteen per cent the use of two or more; in only three per cent three or more.[2]

---

[1] "It is a great matter to observe propriety in these several modes of expression, as also in compound words, strange (or rare) words, and so forth. But the greatest thing by far is to have a command of metaphor. This alone cannot be imparted by another; it is the mark of genius, for to make good metaphors implies an eye for resemblances." "The Poetics of Aristotle" translated by S. H. Butcher, The Macmillan Co., Fourth edition, 1922, p. 87.

[2] Personification is not on the same plane of complexity as metaphor though the line of demarcation is an indefinite one. "White-robed trees" implies the personification of trees but it is metaphorical since it compares snow to white robes.

In the analysis of the phonetic aspects of these short poems, the following facts should be noted: that while in seventy per cent the general iambic pattern found expression, in only eight per cent was regular iambic meter possible; in thirteen per cent the general trochaic pattern and in only two per cent regular trochaic meter; in thirteen per cent the general anapestic pattern and in one per cent regular anapestic meter; and in four per cent the general dactyllic pattern. We do not include irregularities deliberately chosen to enhance a meaning. These figures in general throw light on the comparative difficulty of fitting English syllables to exact phonetic patterns. In terms of regular length of line, the use of pentameter was found in only four per cent and the use of trimeter in nine per cent as opposed to the use of tetrameter in thirty-two per cent and a combination of tetrameter and trimeter in thirty-two per cent. In terms of rhyme, while the use of consecutive and alternate rhymes seem of equal difficulty, being found in the former case in forty per cent and in the latter in forty-one per cent of the poems, skillfully mixed rhymes were found in only eight per cent. In terms of number of rhymes, the use of five or more rhymes was found in 10 per cent; six or more rhymes in 6 per cent. The use of alliteration or the repetition of the initial sound within four accented syllables was found four or more times in seventeen per cent; five or more times in thirteen per cent; six or more times in nine per cent. Assonance or vowal agreement within four accented syllables was found eight or more times within the poem in twenty-two per cent; ten or more times in fourteen per cent; twelve or more times in eight per cent. Consonance or the agreement of consonants was found six or more times in twenty-five per cent; eight or more times in fourteen per cent; twelve or more times in seven per cent.

In judging originality whether of semantic or phonetic pattern, criticism of juvenile verse must be especially guarded, for originality is not a relation between reader and thing written, but between writer and thing written. It is therefore important to know the young writer's past literary experience. Nor can critics dismiss juvenile verse as merely imitative, for so-called imitation involves an organic response, which signifies that the writer is sensitive to the same or similar experiences which moved the original author. That such verse may not survive is an issue apart from the first duty of the teacher or critic which is the recognition of ability. Any deviation

from exact duplication of what has been heard or read may be weighed as a token of originality.

An analysis of a group of poems in which the subject "Snow" has been kept constant reveals that selection tends in the direction of the concrete and the contrasting effect. Simultaneously with the narrowing of choice of descriptive word there is a broadening of the ideational relations. The more complex stages of poem evolution show figures of speech reinforcing one another and the semantic and phonetic aspects of the poem becoming an organic entity comprehended as a unity.

ON BITING VERSUS BARKING CRITICISM

WALTER F. DEARBORN

Harvard University

In the February number of this Journal[1] Professor Spearman charges me with meddling in his affairs—of "barking in" on his dog fight. He complains that, although I make many quotations from his book on "The Abilities of Man," all my "quotations and references come from the smaller and controversial part" of his book and that I have not concerned myself at all with the larger and fundamental part of his work. As a matter of fact, I have not quoted from or made any reference to this book. My book on "Intelligence Tests," to which he refers, was in the press when "The Abilities of Man" was published and the lectures which are reproduced in the book were delivered two years previously. My only quotation from Spearman, as is clearly indicated both in a footnote and in a bibliographical reference, was from his earlier work on "The Nature of Intelligence and the Principles of Cognition." His further remarks are quite irrelevant to the single point of criticism which I have made of him, and equally inaccurate. Either in the confusion and excitement of combat he is uncertain as to the source and nature of his injuries or he is purposely barking off the trail.

In the citations which I made from Spearman's earlier work, he tries to make out that Binet and Simon had no consistent ideas about intelligence, that such ideas as they did have they had abandoned in practice in favor of the two factor theory of Spearman and that "to this complete, however unconscious practical recantation their extraordinary triumph would appear to have really been indebted." This brazen, if belated, effort to steal some of Binet's thunder deserves a bit of ridicule and I remarked: "Thus from his own lips we have it that Spearman had the right theories, and Binet the right methods." My only other comments were to the effect that the methodology of the Binet tests is equally well accounted for by what has been regarded as the diametrically opposed theory of Thorndike and that Binet did have some ideas of his own, preciously like the common factor of Spearman.

If Spearman really has this matter in mind when he finds in my criticism of him "nothing which can be verified by further observation,

[1] Spearman, C.: A Truce to Barking In, *Journal of Educational Psychology*, Vol. XXI, No. 2, pp. 110–111.

or even clarified by further discussion, nothing in a word which makes for any scientific progress," then I should say that this is one way, if a poor one, of admitting the allegations. It is, I suspect, the biting rather than the barking in this and other recent criticism which is disturbing Professor Spearman.

ON THE RELIABILITY OF A STANDARD SCORE

JOHN W. DICKEY

State Normal School, Newark, N. J.

Standard scores made on one test are frequently taken as evidence of standard scores which would occur on another test which measures a quite different function. The reliability of the prediction is known to be a definite function of the reliabilities of the two tests.[1] Moreover, standard scores made on one form of a test are often taken as evidence of the standard scores which would occur on a similar form of the same test. The reliability of this prediction is also known; and is a function of the reliability of the test.[2] What is more interesting is the reliability of the prediction when an obtained standard score is taken as evidence of the best estimated "true" standard score. (The best estimated "true" score is used as the criterion, because the "true" score is never obtained.)

The purpose is here to derive a formula which will measure the reliability of a standard score when it is taken as evidence of the best estimated "true" standard score. The difference between an obtained standard score and the best estimated "true" standard score is a measure of the reliability of the obtained standard score. If an individual were tested with a very large number of similar tests, which measured the same function, the differences between the obtained standard scores and the best estimated "true" standard score would distribute themselves somewhat normally. The standard deviation of the distribution of these errors of estimate measures the reliability of an obtained standard score. The derived formulas will give the standard error, and the probable error of the N errors of estimate for a group; and along with this, they furnish a very close approximation to the standard error, and the probable error of many similar errors of estimate made by a single individual on a very large number of similar tests.

Let,

x_1 = the deviation of a gross score from the mean of the obtained gross-score series,

[1] Kelley, T. L.: *Journal of Educational Psychology*, September, 1923. See also Karl J. Holzinger: *Journal of Educational Pyschology*, January, 1925, for a note on Kelley's proof.

[2] Holzinger, Karl J.: "Statistical Methods for Students in Education." New York, Ginn and Company, 1928, p. 251.

σ_1 = the variability of the obtained gross scores,
$z_1 = x_1/\sigma_1$ = an obtained standard score,
r = the reliability of the test,
x_∞ = the best estimate of a "true" deviation,
σ_∞ = the best estimate of the "true" variability,[1]
$z_\infty = x_\infty/\sigma_\infty$ = the best estimate of a "true" standard score,
ϵ = an error of estimate; that is, $z_1 - z_\infty$.

The error of estimate ϵ is,

$$\epsilon = \frac{x_1}{\sigma_1} - \frac{x_\infty}{\sigma_\infty} \tag{1}$$

Substituting the following equalities,[2]

$$x_\infty = rx_1$$

and[3]

$$\sigma_\infty = \sigma_1\sqrt{r}$$

in equation (1), it becomes

$$\epsilon = \frac{x_1}{\sigma_1} - \frac{x_1\sqrt{r}}{\sigma_1}.$$

Squaring, summing for the population, and dividing by N, there results

$$\frac{\Sigma\epsilon^2}{N} = \frac{\Sigma x_1^2}{N\sigma_1^2} - \frac{2\sqrt{r}\Sigma x_1^2}{N\sigma_1^2} + \frac{r\Sigma x_1^2}{N\sigma_1^2},$$

or

$$\sigma_\epsilon^2 = 1 - 2\sqrt{r} + r,$$

and the required standard error of estimate is

$$\sigma_{(z_1 - z_\infty)} = 1 - \sqrt{r} \tag{2}$$

The reliability of a standard score, when it is taken as evidence of the best estimated "true" standard score, is seen to be a function of the reliability of the test. This conclusion is logically sound. As examples, if $r = 1$, an obtained score is "true"; if $r = 0$, the standard error of estimate would be equal to the standard deviation of the

[1] The best estimate of the "true" variability is used instead of the variability of the best estimated "true" scores, because of the systematic error that would be introduced; that is, an error would be introduced in estimating the "true" scores, which would then be compounded in the calculation of the standard deviation.

[2] Kelley, T. L.: "Statistical Method." New York, Macmillan, 1923, p. 214.

[3] Holzinger, Karl J.: "Statistical Methods for Students in Education." New York, Ginn and Company, 1928, p. 253.

z-scores; and a score would be no more reliable than if it were chosen at random.

Equation (2) in probable error units becomes

$$\mathrm{PE}_{(z_1-z_\infty)} = .6745(1 - \sqrt{r}) \qquad (3)$$

As an example, if an individual's score is 2.0, and $r = .81$, his standard score would become $2.0 \pm .07$. The chances are even that his best estimated "true" standard score will fall between 1.93 and 2.07.

These formulas, (2) and (3), should not be used interchangeably with the formulas,

$$\sigma_{(z_1-z_1)} = \sqrt{1 - r}$$

and

$$\mathrm{PE}_{(z_1-z_1)} = .6745(\sqrt{1 - r}),$$

which yield the reliability of an obtained z-score on one form of a test when it is taken as evidence of a z-score on a similar form of the same test. The error in predicting is always greater in this latter circumstance.

Conclusions

1. The derived formulas yield the reliability of an obtained standard score, if it is taken as evidence of the best estimated "true" standard score.

2. The reliability of a standard score, if it is taken as evidence of the best estimated "true" standard score, is a definite function of the reliability of the test.

3. All standard scores from a given test have equal standard errors, and equal probable errors.

4. The error in predicting the best estimated "true" standard score is always less than the error in predicting the standard score on a similar form of the same test.

THE PROBABLE ERROR OF THE ACCOMPLISHMENT QUOTIENT

C. L. HUFFAKER

University of Oregon

Recently, Odell[1] has given us a comprehensive view of the various measures of achievement relative to capacity. In this study it is shown that the "accomplishment quotient" is the most widely used. The study also points out the need of a measure of reliability of the various devices used. Other writers, notably Kelley and Ruch, have noted the same need. The writer presents below the derivation of a formula to meet this need.

Since the accomplishment quotient is obtained by dividing EQ by IQ it should be treated as a ratio and the formulæ of correlation applying to ratios are the correct ones to use. Holzinger[2] shows that the reliability of a ratio is given by the formula

$$r_{\frac{x_1}{y_1}\frac{x_2}{y_2}} = \frac{r_{x_1x_2}V_{x_1}V_{x_2} + r_{y_1y_2}V_{y_1y_2} - r_{x_1y_2}V_{x_1}V_{y_2} - r_{x_2y_1}V_{x_2}V_{y_1}}{\sqrt{(V^2_{x_1} - 2r_{x_1y_1}V_{x_1}V_{y_1} + V^2_{y_1})(V^2_{x_2} - 2r_{x_2y_2}V_{x_2}V_{y_2} + V^2_{y_2})}} \quad (1)$$

In this formula $r_{x_1x_2}$ and $r_{y_1y_2}$ are reliability coefficients usually expressed as r_{11} and r_{22}. Coefficients such as $r_{x_1y_2}$, $r_{x_2y_1}$, etc., are the various measures of correlation between X and Y or the numerator and denominator of the ratio.

$$V_{x_1} = \frac{\sigma_{x_1}}{M_{x_1}}; \; V_{x_2} = \frac{\sigma_{x_2}}{M_{x_2}} \text{ etc.}$$

In dealing with the accomplishment quotients, X = EQ and Y = IQ. Since we are using reliability coefficients it is necessary that the two measures of X should be comparable, that is have the same means and standard deviations. The same is true of the two measures of Y. In this case we have $M_{x_1} = M_{x_2} = M_1$; $M_{y_1} = M_{y_2} = M_2$; $\sigma_{x_1} = \sigma_{x_2} = \sigma_2$; and $\sigma_{y_1} = \sigma_{y_2} = \sigma_2$.

Therefore

$$V_{x_1} = V_{x_2} = V_1 \text{ and } V_{y_1} = V_{y_2} = V_2.$$

---

[1] Odell, C. W.: A Critical Study of Measures of Achievement Relative to Capacity. *University of Illinois Bulletin* No. 45, March, 1929.

[2] Holzinger, Karl J.: Formulas for the Correlation between Ratios. *Journal of Educational Psychology*, Vol. xiv, September, 1928, p. 346.

We may also assume[1] that $r_{x_1y_1} = r_{x_1y_2} = r_{x_2y_1} = r_{x_2y_2} = r_{xy}$ or r_{12}. Under these conditions Formula (1) reduces to

$$\underset{\bar{y}_1\ \bar{y}_2}{=} \frac{r_{11}V_1^2 + r_{2II}V_2^2 - 2r_{12}V_1V_2}{V_1^2 - 2r_{12}V_1V_2 + V_2^2} \tag{2}$$

Yule[2] gives the formula for the standard deviation of a ratio. Using the same notation as before the standard deviation is given by

$$\sigma_{\frac{x}{y}} = \frac{M_1}{M_2}\sqrt{V_1^2 - 2r_{12}V_1V_2 + V_2^2} \tag{3}$$

The PE of a score is given by the formula[3]

$$\text{PE}_{\text{score}} = .6745\sigma_2\sqrt{1 - r_{11}} \tag{4}$$

Substituting the values obtained from formulas 2 and 3 in 4 and simplifying, we have

$$\text{PE}_{\text{AQ}} = .6745\frac{M_1}{M_2}\sqrt{V_1^2(1 - r_{11}) + V_2^2(1 - r_{2II})} \tag{5}$$

or

$$\text{PE}_{\text{AQ}} = .6745\frac{M_1}{M_2}\sqrt{\left(\frac{\sigma_1}{M_1}\right)^2(1 - r_{11}) + \left(\frac{\sigma_2}{M_2}\right)^2(1 - r_{2II})} \tag{5a}$$

Since the accomplishment quotient technique assumes that EQ's and IQ's are directly comparable, if we assume that

$$\sigma_1 = \sigma_2$$

and

$$M_1 = M_2,$$

then

$$V_1 = V_2 = \frac{\sigma_1}{M_1}.$$

Formula (5) becomes

$$\text{PE}_{\text{AQ}} = .6745\frac{\sigma_1}{M_1}\sqrt{2 - r_{11} - r_{2II}} \tag{6}$$

If as is customary the accomplishment quotient is expressed with two decimal places dropped, it is necessary to multiply the results of Formulæ 5, 5*a* and 6 by 100 in order to secure comparable data.

---

[1] If, as is true in many cases, the reliability coefficients are obtained by finding $r_{\frac{1}{2}\frac{I}{II}}$ and stepping up the result by use of Brown's formula; this is the only possible assumption. For most purposes r_{12} may be regarded as accurate enough for practical uses.

[2] Yule, G. U.: "Introduction to Theory of Statistics." Sixth edition, Charles Griffin, London, 1922, p. 215.

[3] Kelley, T. L.: "Interpretation of Educational Measurements." World Book Co., Yonkers-on-Hudson, 1927, p. 171.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Psychologies of 1930, Edited by Carl Murchison. Worcester, Mass.: Clark University Press. Pp. 497.

This volume continues the series begun with "Psychologies of 1925," the earlier expositions having been re-written and revised and numerous other topics or "schools" included. Half-tone photographs of all the authors are included, and most of the papers are accompanied by a selected bibliography. The editor asserts that in arranging for this volume he has heeded all the helpful criticisms concerning the earlier collection. Such genial suggestibility is not an unmixed virtue, as the contents of the present volume reveal. Instead of nine definitive papers, there are here twenty-five authors, one of them twice represented. The place of several of these in such a volume might easily be questioned and the quality of the papers is very uneven. The book is well printed and errors are few, only seven typographical ones having been noted by the reviewer (pp. 8, 63, 124, 193, 213, 229, 239). The general idea back of the series is an excellent one, and should be continued. The following brief comments will inform the reader of the general character of the volume.

McDougall reiterates the intelligibility and consistency of the "hormic" vocabulary, making out a good case for the second noun only. He still fails to recognize that a desire or a foresight is as natural an event as an itch or a color and that an adequate psychology must confront one as frankly and descriptively as the other.

Brett sketches the "vestiges" of associationism in modern psychology. Actually the vestiges loom large (as in most of the chapters of this book) even when, as in this case, certain important modern tendencies are not represented.

Carr contributes a substantial defence, both historical and dialectical, of the programs of psychologists who are called "functional."

Washburn gives a concise summary of the main features of her ingenious scheme of epiphenomenalism.

Bentley outlines the terminology of a somato-functional psychology which shall be independent of collateral enterprises. Perhaps this very aloofness is what makes the account so quaintly suggestive of scholasticism, in spite of the biological orientation.

Boring sketches the alternatives offered and choices taken by an eclectic program, with "vitality" and "fertility" as its criteria.

Nafe, under a curious caption, writes on "experimental psychologies," concluding that there are no important differences between them.

Three "configurationists" contribute. Köhler indicates in various fields the "tasks" of a *gestalt* psychology, and Koffka applies the principles of stress and organization to the problems of space perception. Sander contends profusely that esthetic preferences for particular patterns require the assumption of "transphenomenal psychic realities" and that the assumption ought to be extended to "the whole realm of psychic reality."

Under "Russian Psychologies" Pavlov succinctly recites the familiar physiological observations of conditioning. Schniermann verbosely describes Bekhterev's reflexology, and Kornilov presents a bizarre intellectual hash under the name "dialectic materialism."

For "Behaviorism," Hunter defends the narrow interest in those visual-tactile objects called men, and for some reason seems to feel a conflict or at least a contrast between this hobby and the larger subject of psychology as a whole. Weiss briefly defends his program, based on the arbitrary distinction between biophysical and biosocial events. Why stop short in the one case and not in the other? Movements actually never "stop short"; all of them can be further traced in neighboring protonelectron aggregates.

Dunlap re-endorses his general principles as previously outlined, and adds certain later ventures—the word *appet* and a revived interest in peripheral tissues as contrasted with the brain.

The genuinely eclectic paper of the series is the brief one by Woodworth on "Dynamic Psychology."

Spearman briefly recapitulates the statistical and the logical features of his activities, professes to tie the two together, and offers the skeleton to psychology as "a school to end schools." He also has a further paper on the various conceptions of normality and abnormality.

"Psychoanalysis" comes in for a short paper by Janet, a longer one by Flügel, and a characteristic one by Adler.

Dewey contributes, under the title "Conduct and Experience," a critical logical analysis of the problem and method of psychology, emphasizing the necessity of recognizing certain "acts" as "mental."

Kelley analyzes "the important steps called for in the study of mental inheritance," and Troland offers a "complete psychophysical theory of motivation." H. L. HOLLINGWORTH.

Barnard College, Columbia University.

Gestalt Psychology, by Wolfgang Köhler. New York: Horace Liveright, 1929. Pp. XII + 403.

The reviewer was impressed with the idea, all the time he was reading this book, that it might have been entitled "Gestalt Psychology, written so you can understand it!" While of course he may be partly under the influence of a euphoristic delusion, the book seems to him to be the clearest statement of the configuration psychology that he has yet seen.

The book begins with an attack upon Behaviorism. In the view of the author the behaviorist is necessarily dependent upon direct experience. But for that matter, so is the physicist! We are told that two physicists can make the same observation on the same galvanometer. How absurd! We have here two entirely different processes in two different organisms. Behaviorists in general do not show an undue interest in epistemology. It is just *one* problem which arouses their interest: How can I know about the direct experiences of others? They entirely neglect a similar problem: How can I prove the existence of a physical world? Of course we believe in this physical world and postulate its existence. The author proposes to make a similar postulate in regard to the direct experiences of other persons.

The scruples of the behaviorists are a serious impediment to the development of the science. Psychology is a very young science, and we need a healthy naïveté which is willing to ignore all epistemological sophistications. We should guard also against a premature effort to secure precise quantitative results. We need at the present time more rough qualitative observations. Our ideal should be Cavendish, who measured the resistances of different materials to an electric current by comparing shocks in his arm. The elaborate quantitative work of Fechner was premature, and perhaps also the contemporary statistical treatment of intelligence tests.

In their reaction from the introspection psychology the behaviorists became "negatively conditioned to much more than the 'harmful stimulus' alone." Gestalt psychology retains the verbal report, but in some other respects is very different from the older introspectionism. The keynote of the new doctrine is organization, a fact which to the reviewer suggests the influence of Hegelianism. Organization is for Köhler a natural and primitive fact. There is a general dynamical interdependence throughout the "field." We see a white object in a shadow as white because of this interdependence, not because we have learned in previous experience that the object is really white. Experience is subordinate to this natural organization. When a well known form is presented in such a way that it is divided between two other well organized but unfamiliar forms, the former is not noticed.

Even animals recognize these forms. Birds will find food under the right flower pot if it is placed so as to stand out from other pots. Distance is in itself unimportant, for the right pot may even be touching some of the others, provided it occupies a strategic position.

Inferences from nonsense syllable experiments are apt to be faulty because of the implication that neutral elements are associated by mere contiguity. However, association is based on organization. Neighborhood in space and time influences association only in so far as it determines organization. We have spontaneous association where the organization is spontaneous. Intentional association takes place where the material is only slightly organized. The author does not feel that "conditioned reflexes" represent any theoretical advance over "associations."

In a final chapter on "Insight," the author reveals his pleasure in a glass of fresh beer on a hot summer day. (He never forgets that he is writing for American readers.) It is not by learning that I know the pleasure to be connected with the taste, not with the spider on the wall or with the size of a certain chair. No, the connection is immediate. The mere ability of the mind to maintain its course through a mental task proves that there is something involved besides habit. The influence of habit alone would mean the speedy dissipation of this purpose in a mass of extraneous associations.

The reader of the book will not look in vain for the old favorites, the chick choosing the darker of two grays and the ape using a stick to reach for a banana. But of these there is rather less than might have been expected.

Although the book seems to the reviewer to be stronger on the destructive than on the constructive side, it is well worth the reading to all those who are struggling in the faith that although they now see through a glass darkly they will some day understand Gestalt psychology.

MELVIN RIGG.
Kenyon College.

A Syllabus Work-book for a Two-level Plan of Instruction in Elementary Psychology, by H. H. Remmers and G. C. Brandenburg. New York: The Macmillian Company, 1930. Pp. VII + 110.

Where classes are large the task of developing initiative, proper attitudes, good study habits, etc., are vastly increased. The lecture device which has accompanied the attempt to instruct large groups of students may come to be a device "by which the contents of the professor's notes get into the note-book of the student without passing through the mind of either." Since the vocal sound and the printed symbol are but devices for stimulating thinking, being different only in the avenue of stimulation, reading may be a device by which the content of the author's mind may serve to generate a memory content of meaningless symbols in the mind of the student, producing no true learning and creating only such attitudes as serve to check future reflective thinking in the field.

A Syllabus Work-book for a Two-level Plan of Instruction in Elementary Psychology represents a timely device for the improvement of instruction. The two-level phase of the plan encourages the student with ability to take hold for himself and investigate further. The opportunity afforded by such an organization is vital to true learning which must of necessity be an active rather than a passive process, for the instructor or the textbook can do no more than stimulate the individual to reorganize his past direct experience. Hence, this book should enable the instructor to do more easily and effectively that which the good instructor has always attempted to do, *i.e.*, develop such attitudes, study habits, concepts, and information as will result in critical thinking and cause the student to continue to develop long after the course is over.

The flexible organization of the work-book provides a minimum amount of definitely assigned work for all with additional work for those who wish to identify themselves with second level procedure. The student is constantly stimulated to effort by the presence of an

objective means of displaying what he has done. By this device it is possible for an instructor to keep some contact with each student in a large class. Thus provision is made for individual differences in interest and ability.

The book contains forty-one lessons with the number of class periods to be given to each suggested by the authors. Each lesson contains definite assignments, the required readings being selected from Gates' "Elementary Psychology" or Dashiell's "Fundamentals of Objective Psychology." Either book may be used as a basic text. Additional readings are selected from such other sources as were considered best. The assignment is followed by a list of questions and suggestions which serve to stimulate thinking and as points of departure for class discussion. Finally, each lesson contains required and optional class preparation in the form of reactions to the materials read.

J. HERBERT BLACKHURST.

Drake University.

The Teaching Unit: A Type Study, by Douglas Waples and Charles A. Stone. New York: D. Appleton and Co., 1929. Pp. X + 205.

The authors of this book state that "The relatively few investigations of problems in teaching method as compared with other types of problems in education may be attributed in part to the lack of suitable techniques for the study of problems in methods" (p. 11). They state further that "The research student too frequently selects a problem adapted to a pre-determined technique. He should instead devise a technique adapted to a problem conspicuously in need of investigation."

In their book on "The Teaching Unit" the authors have made an important step toward the correction of this lack of research in the field of methods. They have made a very painstaking investigation of teaching technique in one very restricted field and have described their methods of investigating the teaching of this one unit so carefully that other investigators should be able to get considerable guidance in carrying out investigations of teaching technique in other units of instruction.

The authors have selected the unit "Positive and Negative Numbers" from junior high school algebra and have devoted the entire book to the analysis of teaching procedures for this one unit. They state that they chose this unit from the field of mathematics because

of its definiteness but they hope by that means to exemplify a method of attack which can be used in studying units which are not so definite.

They have made a detailed analysis of this unit to determine the desired outcomes or objectives, have made diagnostic studies of pupils' difficulties in achieving these various objectives, have collected methods of teaching to remedy these difficulties, and thus formulated a program for teaching the unit "Positive and Negative Numbers."

They give a brief summary (pp. 194–195) of the procedure recommended for the collection of effective methods for teaching a unit in any subject. Briefly stated, the steps are as follows:

1. Select a unit which has demonstrable social value.
2. Record the activities of pupils and teachers while studying this unit.
3. Classify these activities and arrive at tentative objectives.
4. Revise this list of objectives in the light of judgments of representative specialists.
5. Test pupils for the objectives and tabulate the frequency of errors.
6. Analyze and specify efficient methods of performing the operations in which errors occur. Then classify the methods by types.
7. Devise tests for each type of method separately and determine the seriousness of difficulties by the number of pupils unable to apply each method.
8. Select teaching methods for removing the difficulties of each type and apply the method to the pupils confronting each type of difficulty. Test again to determine what difficulties persist.
9. Interview persons in occupations unrelated to school teaching to obtain new methods of overcoming the learning difficulties.
10. Continue to experiment with the teaching methods obtained from all sources until methods are found which remove the difficulties presented by the given unit as taught by the given staff in the given school.

Because of the fact that the book is a type study and not a comprehensive textbook, its value will be mainly for persons interested in research in the field of methods. Its value here will be very high. The suggestions are so concrete, however, that a number of more than ordinarily interested classroom teachers and supervisors will be able to follow the path that has been mapped out and thus accomplish significant improvement in the teaching of other units and other subjects.

C. C. Crawford.

University of Southern California.